FOSTERING INCIDENTAL VOCABULARY UPTAKE

FROM AUDIO-VISUAL MATERIALS:

THE ROLE OF TEXT COMPREHENSION

By

Nguyễn Chí Đức

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

----2017----

Abstract

This research project explores various factors that may influence the rate of incidental foreign/second language (L2) vocabulary acquisition from audio-visual materials, with a special focus on procedures that enhance learners' comprehension of these input materials. Informed by relevant theories and research findings in the fields of L2 listening comprehension and incidental vocabulary acquisition, I investigate the effects of having learners (a) view a TED Talks video twice rather than once, (b) sum up the content of the video before watching it a second time, (c) watch TED Talks videos on the same subject in order to increase familiarity with that subject, and (d) exchange summaries of TED Talks videos with peers so as to assist each other's subsequent processing of those videos. As these interventions are all deemed to facilitate L2 listening comprehension, they are also expected to create favourable conditions for incidental vocabulary uptake to occur.

The effects on incidental vocabulary acquisition of the above interventions were gauged in a series of classroom experiments with Vietnamese EFL learners. Although vocabulary uptake was generally far from spectacular, all of the tested procedures were found to result in statistically significant vocabulary gains. The insertion of the output tasks (i.e., the summary activities) was particularly useful. First, they helped to enhance the learners' text comprehension. Second, they created opportunities for the learners to use newly met words and thus consolidate their knowledge of these lexical items. A thread through the experimental data is the strong association between the learners' vocabulary uptake and their comprehension of the input content.

The findings from this research project are consistent with several established notions, models and theories in the fields, including Ausubel's Advance Organizer (1978), Hulstijn and Laufer's Involvement Load Hypothesis (2001), Krashen's Input Hypothesis (1985), Nation's Vocabulary Generation (2013), Swain's Output Hypothesis (2005), and Wittrock's Model of Generative Teaching of Comprehension (1991). However, there are also findings that go beyond the core tenets of these, and that can further our understanding of how learners process new lexical items in meaning-focused input and output tasks.

Regarding pedagogical implications, this research project confirms that fostering L2 listening comprehension creates favourable conditions for incidental vocabulary acquisition to happen, and that the aforementioned classroom procedures are facilitative in this regard, albeit to different degrees.

Table of Content

Abstract	i
Table of Content	ii
List of Tables, Diagrams and Abbreviations	vii
Acknowledgement	x
INTRODUCTION	1
CHAPTER 1: LITERATURE REVIEW	4
Vocabulary Knowledge and Incidental Vocabulary Acquisition	4
A Generally Small Vocabulary Gain from Listening Input	6
Role of Text Comprehension in Incidental Vocabulary Acquisition from Listening Input	10
Intervention 1: Using Video rather than Audio Materials as Listening Input..	13
Intervention 2: Viewing a Video Twice rather than Once	17
Intervention 3: Adding an Output Task between these Two Viewing Times .	20
Intervention 4: Viewing Three Videos on Similar Topics	28
Intervention 5: Adding an Output Task into this Narrow Viewing Procedure	31
Summary	35
CHAPTER 2: METHODOLOGY	38
Research Aims, Research Questions and Research Designs	38
Research Participants	41
Input Materials and Potential Target Words	42
Input Materials	42
Potential Target Words	46

Dependent Measures in Study 1, 2, 3 and 7	48
Text Comprehension Test	49
Selecting and operationalizing the test construct	49
Designing the test	50
Validating the test	51
Vocabulary Knowledge Tests	58
Selecting and operationalizing the test construct	58
Designing the tests	58
Validating the tests	61
Dependent Measures in Study 1, 2, 3 and 7	64
Text Comprehension Test	64
Designing the test	64
Validating the test	65
Vocabulary Knowledge Tests	69
Designing the tests	69
Validating the tests	69
Procedure of Data Collection	71
Data Analysis	74
CHAPTER 3: RESULTS	77
Study 1: OV	77
Word-form Gain	78
Word-meaning Gain	79
Association between Vocabulary Gain and Text Comprehension	80

Associations between Vocabulary Gain and other Factors of Interest	81
Summary	82
Study 2: RV	82
Word-form Gain	82
Word-meaning Gain	84
Association between Vocabulary Gain and Text Comprehension	85
Associations between Vocabulary Gain and other Factors of Interest	85
Summary	86
Study 3: RVS	86
Word-form Gain	87
Word-meaning Gain	88
Association between Vocabulary Gain and Text Comprehension	89
Associations between Vocabulary Gain and other Factors of Interest	89
Summary	90
Study 4: NV	90
Word-form Gain from the Longest Video	91
Word-meaning Gain from the Longest Video	92
Word-form Gain from all Three Videos	93
Word-meaning Gain from all Three Videos	94
Association between Vocabulary Gain and Text Comprehension	95
Associations between Vocabulary Gain and other Factors of Interest	96
Summary	97
Study 5: NVS	97

Word-form Gain	98
Word-meaning Gain	99
Association between Vocabulary Gain and Text Comprehension	100
Associations between Vocabulary Gain and other Factors of Interest	100
Summary	101
Study 6: A Comparison of Effect Sizes on Vocabulary Gain	101
A Comparison of Effect Sizes on Vocabulary Gain.....	101
Explanatory Factors for the Difference in these Effect Sizes	103
Vocabulary gains in the OV, the RV and the RVS condition	103
Vocabulary gains in the NV and the NVS condition	107
Vocabulary gains in the OV and the NV condition	111
Summary	114
Study 7: A Qualitative View into the Benefits of the Output Tasks for	
Vocabulary Gain	114
Benefits of the Monologue Summary Activity in the RVS Condition	115
Quantitative results	115
Qualitative results	116
Benefits of the Interactional Summary Activity in the NVS Condition	119
Quantitative results	119
Qualitative results	119
An Overall Summary of the Research Results	123
CHAPTER 4: DISCUSSION AND IMPLICATIONS	125
Discussion	125

Study 1: OV	125
Study 2: RV	128
Study 3: RVS	131
Study 4: NV	135
Study 5: NVS	138
Study 7: A Qualitative View into the Benefits of the Output Tasks for Vocabulary Gain	141
Summary	147
Implications	149
Implications for Instructional Practice	149
Pedagogical values of the interventions	149
Selecting suitable TED Talks	151
Designing classroom procedures	153
Implications for Future Research	158
CONCLUSION	163
References	166
Appendices	179

List of Tables, Diagrams and Abbreviations

Tables

Table 1 Nation's construct of vocabulary knowledge	5
Table 2 Suggested interventions in this research project	35
Table 3 Research aims and research questions	38
Table 4 Potential target words for Study 1, 2, 3 and 7	46
Table 5 Potential target words for Study 4, 5 and 7	47
Table 6 Read's dimensions of L2 vocabulary assessment	58
Table 7 Overall procedure of data collection	71
Table 8 OV condition: Word-form gain	78
Table 9 OV condition: Word-meaning gain	79
Table 10 OV condition: Associations between vocabulary gain and other factors ...	81
Table 11 RV condition: Word-form gain.....	83
Table 12 RV condition: Word-meaning gain.....	84
Table 13 RV condition: Associations between vocabulary gain and other factors ...	85
Table 14 RVS condition: Word-form gain	87
Table 15 RVS condition: Word-meaning gain	88
Table 16 RVS condition: Associations between vocabulary gain and other factors .	89
Table 17 NV condition: Word-form gain from the longest video	91
Table 18 NV condition: Word-meaning gain from the longest video	92
Table 19 NV condition: Word-form gain from all three videos	93
Table 20 NV condition: Word-meaning gain from all three videos	95

Table 21 NV condition: Associations between vocabulary gain and other factors ...	96
Table 22 NVS condition: Word-form gain	98
Table 23 NVS condition: Word-meaning gain	99
Table 24 NVS condition: Associations between vocabulary gain and other factors .	100
Table 25 Association between vocabulary gain and summary quality	109
Table 26 Text comprehension in the OV and the NV condition	112
Table 27 Vocabulary gain in the NV condition	113
Table 28 RVS condition: Text comprehension and vocabulary gain	115
Table 29 NVS condition: Text comprehension and vocabulary gain	119
Table 30 An overall summary of the research results	123

Diagrams

Figure 1 Rasch Model's person-item map on comprehension test 1	53
Figure 2 Rasch Model's person-item map on comprehension test 2	65
Graph 1 Effect sizes on word-form acquisition	102
Graph 2 Effect sizes on word-form retention	102
Graph 3 Effect sizes on word-meaning acquisition	102
Graph 4 Effect sizes on word-meaning retention	102
Plot 1 Text comprehension in the OV, the RV and the RVS condition	103
Plot 2 Text comprehension in the NV and the NVS condition	107
Plot 3 Text comprehension in the OV and the NV condition	112

Abbreviations

CEFR: Common European Framework of Reference for Languages

EFL: English as a Foreign Language

ESL: English as a Second Language

IELTS: International English Language Testing System

L1: First language

L2: Foreign or second language

OV: One-time Viewing

RV: Repeated Viewing

RVS: Repeated Viewing plus Summary

NV: Narrow Viewing

NVS: Narrow Viewing plus Summary

TED: Technology-Entertainment-Design

TOEFL: Test of English as a Foreign Language

TV: Television

Acknowledgement

I used to be a person with low self-esteem. I did not believe that I had enough ability to complete any big tasks like this research project. It was my supervisors Associate Professor Frank Boers and Associate Professor Jonathan Newton that helped me to start to believe in myself again, seeded and nurtured my love for research, leaned me against their scaffoldings when my research ability was still too weak to stand on its own feet, and then let go when they felt that I was able to cope with a given research task myself. With their academic guidance and spiritual support, I have seen myself getting academically matured year by year. I would like to take this opportunity to thank them for all what they, regardless of day or night, have done for me.

I also would like to thank Professor Paul Nation (Victoria University of Wellington, New Zealand) and Professor Stuart Webb (University of Western Ontario, Canada) for building the foundation for this research project and lending me a helping hand whenever I need it.

I am deeply grateful to all teachers and students in University of Languages and International Studies, Vietnam and Victoria University of Wellington, New Zealand who took part in this research project either as independent coders/raters or research participants. I do hope that this thesis would give them some useful implications for their English language teaching and learning in return for their great contribution to this research project.

The biggest thank should be given to my mum who was the first person to encourage me to pursue this PhD course, cried with me when I got stuck in my research process, smiled with me when I made even a small progress, and was patiently waiting for me to come back home. She always believed that I could complete this course even though I myself did not believe in it sometimes. I would like to use this thesis as a big present to thank her for believing, loving and caring for me.

INTRODUCTION

An important added value of a meaning-focused foreign/second language (L2) listening activity is its provision of opportunities for incidental vocabulary uptake (R. Ellis, 1999; Nation, 2013; van Zeeland, 2013). In fact, previous research in this field has shown that L2 learners are able to pick up new lexical knowledge from comprehension-based activities with both audio and audio-visual input materials (e.g., Brown, Waring, & Donkaewbua, 2008; R. Ellis, Tanaka, & Yamazaki, 1994; Sawada, 2009; Rodgers, 2013; Vidal, 2003, 2011). However, the vocabulary gain from this source tends to be very limited, especially when authentic input materials are brought into the language classroom without any modification (e.g., Rodgers, 2013). This small vocabulary gain can be attributed to a wide range of factors, such as: the challenging nature of real-time listening, the limited capacity of human working memory, or learners' unfamiliarity with L2 lexical segmentation and recognition in spoken discourse (van Zeeland, 2013). This therefore suggests that incidental vocabulary acquisition from meaning-focused L2 listening activities should not be taken for granted or, in other words, L2 teachers and course designers do need to resort to pedagogical interventions to enlarge the size of vocabulary gain from this source.

A pertinent question here is what type of interventions is conducive to incidental L2 vocabulary uptake from meaning-focused listening activities. One of the answers may lie in the instructional practices that facilitate learners' L2 listening comprehension. This is because, when their L2 listening comprehension is cognitively catered for, learners will have more mental resources to process novel words in listening input (Boers, Eyckmans, & Godfroid, 2007). An adequate level of text comprehension is also believed to enable them to make good use of available contextual clues for their lexical form-meaning mappings (Haastrup, 1991; Nation, 2013; Stæhr, 2009). With this view in mind, I have used relevant theories and research findings in the fields of L2 listening comprehension and incidental vocabulary acquisition to construct five interventions in this research project, as follows: (1) using video rather than audio materials as listening input, (2) having L2 learners view a video twice rather than once, (3) having them view a video, make an oral summary of the content, and then view this video again, (4) having them view a set of three videos on similar topics, and (5) having them view either the first or the second half of this three-video set, share the content with a classmate who has not viewed the same half yet in an interactional summary activity, and then move on with the other half.

In Intervention 1, text comprehension may be fostered thanks to the presence of visual support in the given input material (Baltova, 1994; Jones & Jan, 2002; Parry & Merredith, 1987; Sadoski, 2005; Shin, 1998; Wagner, 2010). In Intervention 2, text comprehension is further aided by the second viewing, where learners can verify their previous interpretation of the input content and add information that they have missed during their first viewing (Bern, 1995; Brown, 2007; Cervantes & Gainer, 1992; Chang & Read, 2006; Dupuy, 1999; Krashen, 1996; Lund, 1991; Saika, 2009). In Intervention 3, the incorporation of an oral summary activity between two times of viewing the same video might help learners to reprocess their perceived information from the previous viewing, and then get prepared for the second viewing (Doctorow, Wittrock & Marks, 1978; Gambrell, Koskinen, & Kapinus, 1991; Wittrock, 1991; Wittrock & Kelly, 1984). Intervention 4 uses a set of three videos on similar topics in order to increase learners' topic familiarity and lexical coverage of the input materials when they move on from one input text to another, which, in turn, facilitates their text comprehension (Bonk, 2000; Krashen's 1996; Leeser, 1998; Long, 1990; Rodgers & Webb, 2011; Sadighi & Zare, 2006; Schmidt-Rinehart, 1994; Schmitt & Carter, 2000; Stæhr, 2009; Zidong, 1998). This text comprehension is expected to be further supported by the interactional summary activity in Intervention 5, where learners are required to summarize the content of the video(s) that they have viewed to a classmate, and listen to this classmate's briefing about the content of the materials that they are about to watch. Receiving a summary of the video(s) before viewing may give learners an advance organizer of upcoming information and therefore foster their listening comprehension (Ausubel, 1960, 1978; Herron, 1994, Herron et al., 1995; Jafari & Hashim, 2012). As these interventions are all deemed to facilitate L2 listening comprehension, they are also expected to create favourable conditions for incidental vocabulary acquisition to occur. In addition, the monologue summary activity in Intervention 3 and the interactional summary activity in Intervention 5 are likely to stimulate incidental vocabulary acquisition in accordance with Hulstijn and Laufer's Involvement Load Hypothesis (2001), Long's Interaction Hypothesis (1996), Nation's Vocabulary Generation (2013) and Swain's Output Hypothesis (2005).

The actual effects on the size of vocabulary gain of these interventions were measured through five major classroom experiments which are henceforth referred to as Study 1 to Study 5 in this research project. Through these studies, I sought the answers to three major research questions:

- A. *To what extent can L2 learners pick up new vocabulary knowledge from each intervention above?*
- B. *Can the degree of text comprehension account for the size of vocabulary gain in each intervention and the difference in the size of vocabulary gains across the interventions?*
- C. *Are there any other factors in the interventions, the input materials or the learners that can also help to explain the size of vocabulary gain in each intervention and the difference in the size of vocabulary gains across the interventions?*

The present thesis includes four main chapters. In the Literature Review Chapter, I shall describe the theoretical grounding and the empirical research that motivated the five aforementioned interventions. In the Methodology Chapter, I will specify the research aims, research questions, research designs, research participants, input materials, target words, dependent measures, and procedures of collecting and analysing data for the five studies. Also in this chapter, I will explain why, in addition to these five studies, two additional investigations including one of a more qualitative nature – Study 6 and 7 – are deemed necessary and have been included in this research project. The findings from these seven studies will be reported in detail in the Results Chapter. In the Discussion and Implications Chapter, I will first identify the key factors that account for the size of vocabulary gain in each intervention and also the difference in the size of vocabulary gains across the interventions. These key factors will then be related back to the relevant theories and research findings that I have used as a foundation to build the interventions upon. Finally, implications and recommendations for instructional practice and for future research will be drawn from the research findings and methodological aspects of the project.

Through this research project, I myself would like to understand more about (a) how L2 learners process new lexical items in meaning-focused listening input and listening-based output tasks, (b) whether this processing leads to incidental vocabulary uptake, and (c) whether the size of this vocabulary uptake is associated with the degree of text comprehension and other input-related and learner-related factors. This understanding will help me as a L2 teacher and course designer to develop meaning-focused listening lessons that foster not only text comprehension but also incidental vocabulary uptake. It is also expected that this research project will provide some useful information for my colleagues in my teaching context in Vietnam.

CHAPTER 1: LITERATURE REVIEW

In this chapter, I shall first describe the way I understand and operationalize two key concepts “*vocabulary knowledge*” and “*incidental vocabulary acquisition*” in this research project (Section 1). This is followed by a brief review of previous research that has measured the size of incidental L2 vocabulary uptake from listening input (Section 2). The result from this review indicates that incidental L2 vocabulary acquisition from listening input is generally limited. Therefore, pedagogical interventions are needed to enhance the size of vocabulary gain from this source. One way to do so is to design pedagogical interventions that foster text comprehension. In this way, learners are expected to have more mental resources to process novel words in listening input (Section 3). With this in mind, I have suggested five different interventions that will be put to the test in Study 1 to Study 5 in this research project: (a) using video rather than audio materials as listening input, (b) having learners view a video twice, (c) adding an oral summary activity between two times of viewing the same video, (d) having learners view a set of three videos on similar topics, and (e) adding an interactional summary activity into the procedure of viewing this three-video set. From Section 4 to Section 8, I will present the theoretical and empirical ground in the fields of L2 listening comprehension and incidental vocabulary acquisition to support my suggested interventions. The final section will summarize the key content of this chapter and explain why Study 6 and 7 are also required in this research project.

Vocabulary Knowledge and Incidental Vocabulary Acquisition

As this research project aims to examine incidental L2 vocabulary uptake from listening input, let me start this chapter by clarifying what is subsumed under two key concepts “*vocabulary knowledge*” and “*incidental vocabulary acquisition*”. In what follows, I shall also describe the way I operationalize these two concepts in my upcoming classroom experiments.

There is a consensus in this field that vocabulary knowledge includes more than merely form-meaning relationships (e.g., Aitchison, 1994; Nation, 1990, 2001, 2013; Richards, 1976). For example, Aitchison (1994) suggested that knowing a word also means being able to build up a semantic map between this word and its wide range of associates, such as synonyms, antonyms or hyponyms. According to Richards (1976), vocabulary knowledge includes, among many other aspects, the knowledge about the grammatical function that a word plays in a particular context and the socio-cultural context(s) in which

this word can or should not be used. To date, Nation's construct of vocabulary knowledge (2013, see Table 1 below) has been considered the most comprehensive and useful for both pedagogy and research practice (Daller, Milton, & Treffers-Daller, 2007, p.4; Read, 2000, pp.26-27). Therefore, this construct will be used to define vocabulary knowledge in the present research project. In this construct, vocabulary knowledge comprises three big domains: form, meaning, and use. In the *form* domain, there are three aspects: spoken word form, written word form, and word parts. In the *meaning* domain, there are another three aspects: form and meaning connection, concept and referents, and word associations. The *use* domain includes the last three aspects: grammatical functions, word collocations, and constraints on word use. Each of these aspects is further divided into two dimensions: receptive knowledge and productive knowledge. The aspects are presented in Table 1:

Table 1: Nation's construct of vocabulary knowledge (2013, 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 the 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 word 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	R	Where, when, and how often would we expect to meet this words?
		P	Where, when, and how often can we use this word?

R = Receptive knowledge; P = Productive knowledge

From an assessment perspective, the only problem with this construct is the challenge of designing a test that measures every single aspect of vocabulary knowledge for a large number of lexical items (Meara, 1996; Read, 2000, p.27). Therefore, it is worth narrowing

down the scope of such a measure to only one or two most representative aspects of this construct. Research investigating incidental vocabulary acquisition from both reading (Chen & Truscott, 2010; Webb, 2007) and listening input (Brown et al., 2008; Horst, Cobb, & Meara, 1998) indicates that L2 learners are able to gain many other aspects of lexical knowledge apart from form-meaning connection. However, this connection is still regarded as the most important among various aspects of vocabulary knowledge (Laufer & Goldstein, 2004; Nation, 2001, 2013). Therefore, it seems justified to use this aspect to operationalize the construct of vocabulary knowledge in the present research project.

Scholars in this field also agree that vocabulary acquisition is considered incidental when learners pick up new vocabulary knowledge as a by-product of a meaning-focused input or output activity (R. Ellis, 1999; Hulstijn, 2003; Nation, 2013; Schmitt, 2010). This learning process is different from intentional vocabulary learning where learners deliberately internalize a new lexical item into their interlanguage system. In classroom experiments as in the case of my research project, incidental vocabulary acquisition is often operationalized by engaging learners in a meaning-focused input task (e.g., listening to a short story for the purpose of text comprehension), which may be followed by a text-based output task (e.g., summarizing the story content), and then having them sit a surprise test that measures their knowledge of unfamiliar words occurring in the given input materials (see Brown et al. (2008), de la Fuente (2002), and Vidal (2003, 2011) for examples). When the presence of this vocabulary post-test has not been announced prior to the input stage of the experiment, vocabulary gain attested in this test can be attributed to incidental learning (Hulstijn, 2003).

So far, I have clarified the conceptualization and operationalization of vocabulary knowledge and incidental vocabulary acquisition in this research project. Now let me move on to review what previous research in this field has found about the size of incidental L2 vocabulary uptake from listening input.

A Generally Small Vocabulary Gain from Listening Input

Previous research has showed that L2 learners are able to pick up new vocabulary knowledge from listening input without any other support (henceforth referred to as “*listening input only*” for short). However, vocabulary gain from this source is generally found to be small, especially as compared to the gain made from reading input. In addition, the size of this vocabulary gain can be influenced by a wide range of factors.

Brown et al. (2008), for example, compared the size of incidental L2 vocabulary uptake across three learning conditions. One of these conditions used listening input only, which is therefore relevant to the purpose of the current review. In this condition, 35 learners of English as a foreign language (EFL) were required to listen to the audio recordings of three graded readers for the purpose of text comprehension. Their vocabulary gain was measured by a word-meaning recall test and a word-meaning recognition test, which were implemented immediately at the end of the experiment, one week, and then three months later. On the word-meaning recall test, these learners were found to obtain a mean gain of 0.56 (*SD* 1.13, 2%), 0.74 (*SD* 1.12, 2.62%) and 0.37 (*SD* 0.72, 1.32%) out of 28 target words right after the experiment, one week, and three months later, respectively. On the word-meaning recognition measure, they made a mean gain of 8.20 (*SD* 2.82, 29.29%) target words in the immediate post-test, 9.06 (*SD* 2.65, 32.36%) target words in the one-week delayed post-test, and 10.09 (*SD* 2.72, 36.04%) target words in the three-month delayed post-test. It is clear from these results that the size of vocabulary gain in this condition heavily depended on the type of vocabulary measures. In addition, the lexical uptake measured on the word-meaning recognition test was also reported to be positively correlated with the frequency of word occurrences in the input materials. This correlation, however, fell short of the required significance level in the case of the word-meaning recall test, which Brown et al. attributed to the floor effect of such a small gain size on this measure.

R. Ellis, Tanaka, and Yamazaki (1994) also compared the size of incidental L2 vocabulary gain across three learning conditions. One of these used listening input only. Therefore, the vocabulary gain found in this condition is also relevant to the purpose of the current review. In this condition, 43 EFL learners listened to a native speaker giving 16 directions to put 19 kitchen items (which were also the target words in this experiment) in their corresponding places in a kitchen map. Their vocabulary gain was gauged by a word-meaning recall test that was carried out twice - two days and then one month after the experiment – and a form-meaning matching test that was implemented two months and a half later. In these three post-tests, the learners were found to obtain a mean gain of 2.02 (10.63%), 2.53 (13.32%), and 4.05 (21.32%) out of the 19 target words respectively (the corresponding standard deviations were not reported).

Van Zeeland and Schmitt (2013) also examined the effect of listening input on incidental vocabulary uptake. In this study, 30 learners of English as a second language (ESL) were required to listen to four passages for the purpose of text comprehension. Their

vocabulary gain was determined by means of a word-form recognition test, a part-of-speech recognition test, and a word-meaning recall test, which were conducted immediately at the end of the experiment and then two weeks later. In these immediate and delayed post-tests, out of 24 target words, the learners made a mean gain of 11 (45.83%) and 6 (25%) on the word-form recognition measure, 8.10 (33.75%) and 5.90 (24.58%) on the part-of-speech recognition measure, and 2.05 (8.54%) and 1.80 (7.50%) on the word-meaning recall measure, respectively. Their lexical uptake was found to be associated with the frequency of word occurrences, the concreteness of word meaning, and the grammatical functions of the target words in the given input materials. Specifically, the more often the target words were encountered in the input materials, the more likely they were picked up. The target words with concrete meaning proved to be easier for uptake than those with abstract meaning. The difficulty level of acquisition increased from nouns to verbs and then to adjectives.

It is clear from these studies that L2 learners are able to incidentally acquire new lexical knowledge from listening input only. However, the size of vocabulary gain, especially in the case of word-meaning gain, is relatively small. We might have a better view of how limited this lexical uptake is when it is put in comparison with the gain made from reading input. Brown et al. (2008), for instance, compared the size of incidental L2 vocabulary uptake from listening input (as already reviewed above) with that made from reading input. On the word-meaning recall test, the learners in the reading condition ($N = 35$) were found to make a mean gain of 4.10 (SD 4.02, 15%), 2.34 (SD 2.39, 8.54%) and 0.97 (SD 1.47, 3.46%) out of 28 target words right after the experiment, one week, and then three months later, respectively. On the word-meaning recognition, their gains were 12.54 (SD 5.03, 44.79%), 12.46 (SD 4.25, 44.50%) and 11.37 (SD 3.10, 40.61%) out of the 28 target words, respectively. On both measures, these learners scored significantly higher than those in the listening condition, with $t = 5.66$ ($p < .05$) and $t = 6.96$ ($p < .05$), respectively.

It also needs to be acknowledged that these studies all have methodological limitations. First, neither of these studies included a control group in their research designs to gauge the effects of test-taking and chance learning of the target words outside of the experiments. Given the considerable intervals between their post-tests, this inclusion would have been useful. The studies by Brown et al. (2008) and R. Ellis et al. (1994) used the written form of the target words as the test prompt in their vocabulary measures even though their learners were not exposed to this form in the given input materials. This apparently creates a measurement bias in the case of the study by Brown et al. (2008), where the

vocabulary gain from listening input was compared with that from reading input. From the provided written form of the target words as the test prompt in the vocabulary post-tests, the learners in the reading condition could recognize and recall the meaning of more target words than those in the listening condition probably because they had encountered this form in their input materials while those in the listening group had not. Previous research has found a significant difference in the size of vocabulary gains when the presentation mode of the test prompt was either congruent or not congruent with that of the input text (e.g., d'Ydewalle & Van de Poel, 1999; Sydorenko, 2010). This limitation was partly removed in the study by van Zeeland and Schmitt (2013), which used the spoken form of the target words as the test prompt in two out of the three vocabulary measures (i.e., the word-form recognition test and the part-of-speech recognition test). Another merit of this study is its replacement of the target words in the listening input with non-words. As a result, no pre-test of these lexical items was required and therefore the learners were not alerted to the purpose of the experiment. However, the use of non-words and manipulation of the frequency at which these words occurred in the input materials resulted in a substantial modification of the input materials. This raises a concern about input authenticity, especially regarding whether these non-words would behave in the same way as real words do. Given these limitations, more research into incidental L2 vocabulary uptake from listening input is still welcome (van Zeeland, 2013). Therefore, I also carry out this research project with an aim to remove such limitations.

Despite their methodological limitations, the studies reviewed above have provided clear evidence for incidental L2 vocabulary acquisition from listening input. However, the size of vocabulary gain from this source is generally small, especially as compared to the gain made from reading input. In addition, this lexical uptake is also found to be moderated by a wide variety of factors, such as: word-related factors (part of speech and concreteness of word meaning), input-related factors (frequency of word occurrences in input materials), and learner-related factors (their text comprehension, their prior vocabulary knowledge or general L2 competence). This therefore suggests that incidental L2 vocabulary acquisition from listening input cannot be taken for granted or, in other words, pedagogical interventions are needed to enhance the size of vocabulary gain from this source. A pertinent question here is what type of pedagogical interventions has the potential to increase vocabulary gains from listening input. One of the answers may lie in the pedagogical interventions that facilitate L2 listening comprehension. In the following section, I shall discuss why fostering L2 listening

comprehension can create favourable conditions for incidental vocabulary acquisition to occur.

Role of Text Comprehension in Incidental Vocabulary Acquisition from Listening Input

For incidental vocabulary uptake from a meaning-focused listening task to occur, learners need to direct their attention to unfamiliar words in input materials. In his Noticing Hypothesis, Schmidt (1990, 1994, 1995, 2001) described this attention as a necessary (though not sufficient) condition to convert input to uptake. To be more specific, he posited that a new language code (e.g., a novel word) in a given input text can only be converted to uptake when learners notice it, and build a mental representation of this code in their working memory which may then undergo further mental processing. In an attempt to specify the roles of noticing in the conversion of input to uptake, Robinson (1995, 2003) developed a model of working memory in which noticing is conceived of as a combination of two functions – detection and rehearsal. While the former mentally registers a new language code into short-term memory, the latter first uses subvocal rehearsal mechanisms to maintain this code in working memory longer and then conceptually map it to relevant schemata in long-term memory.

However, the chances for this noticing to happen during a meaning-focused listening task are extremely limited. First, this is due to the limited capacity of human working memory (Cowan, 1995; Robinson, 2003; Schmidt, 1990). According to Oberauer and Kliegl (2006), a typical duration that human short-term memory can temporarily hold a manageable chunk of new information is about 18 seconds. In this very short span of time, L2 listeners need to receive and store incoming information in a matter of one or two seconds in their sensory memory (also known as the perception phase), rapidly construct a mental representation of this information in about 15 seconds (normally referred to as the parsing phase), and then immediately transmit this representation into long-term memory (where this representation is ready for the utilization phase) so that their working memory is freed up to process new incoming information (Brown, 1995; Field, 2008; Flowerdew & Miller, 2005; Lynch, 2009; Rost, 2013). Given this limited capacity of working memory, together with their currently yet-to-develop L2 proficiency, learners tend to forget what is heard in trying to keep up with incoming information or busily interpret the received input at the expense of the next part (Goh, 2000). It is obvious from the description above that L2 listening comprehension is already a mentally demanding process for learners. Therefore, when

learners have invested most of their working memory capacity for the purpose of text comprehension, their remaining mental resources are not sufficient to deal with unfamiliar words in the given listening materials.

In addition, learners often have difficulty with their L2 lexical segmentation and recognition when they need to deal with spoken discourse (Vandergrift, 2004; van Zeeland, 2013). According to Vidal (2003, 2011), as learners fail to discern a novel word in connected speech, they cannot notice the presence of this word in the input materials, let alone make use of contextual factors that are often found to facilitate incidental vocabulary acquisition, such as word repetition or the presence of verbal and pictorial clues for word-meaning inferencing. In case they do recognize the presence of this word in the given input, the nature of real-time listening constitutes another impediment to their lexical form-meaning mappings. This is because this type of listening does not allow them to stop the flow of input presentation at any particular lexical item (Montero Perez, Noortgate, & Desmet, 2013) or revisit past passages to deploy contextual clues for their interpretation of the word meaning (Boers, Eyckmans, & Godfroid, 2007).

The discussion above suggests that when L2 listening comprehension is supported, learners will have more mental resources to deal with unfamiliar words in input materials. If so, the likelihood of incidental vocabulary acquisition is improved. In fact, Krashen (1980, 1983, 1985) described the role of text comprehension in L2 acquisition in his Comprehensible Input Hypothesis. Four core tenets in this hypothesis are as follows. First, for L2 acquisition (including incidental vocabulary uptake) to take place, learners need to have comprehensible input. Second, this input must be at a stage which is slightly beyond their current level of L2 proficiency ($i + 1$). Third, when learners have such comprehensible input and their affective state are facilitative of learning (e.g., high motivation and low anxiety), language acquisition can occur, even without their conscious effort. Finally, L2 production (output) as such does not aid acquisition or, to be more precise, it is merely a result of acquisition. For the last three decades, this hypothesis has been criticized for overestimating the role of comprehensible input as a source of language acquisition and for neglecting the roles of cognitive factors (e.g., noticing and/or awareness), interaction and output in this learning process (e.g., R. Ellis, 1991, 1999; Mackey, 2012; Robinson, 2003; Schmidt, 1990, 1993, 2001). However, it should be acknowledged that this hypothesis does point to some important conditions for incidental vocabulary uptake. First, “ $i + 1$ ” input provides learners with opportunities to encounter unfamiliar words in input materials and

therefore to recognize lexical gaps in their interlanguage system. This recognition may prompt them to fill these gaps when they are exposed to L2 input again (Krashen, 1985). At the same time, comprehensible input allows learners to make good use of contextual clues for their word-meaning inferencing (Haastrup, 1991; Nation, 2013; Stæhr, 2009). Therefore, instead of criticizing this hypothesis as a whole, Long (1983) highly valued Krashen's argument about the importance of comprehensible input in language acquisition and modified this hypothesis into a weak version of Krashen's Input Hypothesis where comprehensible input is still held to be a necessary, though insufficient, condition for language acquisition to occur.

Previous research has also provided supporting evidence for the role of text comprehension in incidental vocabulary acquisition from listening input. Sawada (2009), for example, examined the extent to which 116 EFL learners incidentally picked up the meaning of 45 unfamiliar words from listening to nine TOEFL¹ listening passages. These learners were found to obtain a significant vocabulary gain after their listening procedure and the size of their vocabulary gains was positively correlated with the degree of their text comprehension. Vidal (2011) also measured the size of incidental L2 vocabulary gain, but from reading or listening to three academic texts. In this study, 192 ESL learners were assigned to either the reading or the listening condition. Their text comprehension was gauged by means of a true-false test and a cloze test. Their vocabulary gain was determined by a self-report test using a modified version of the Vocabulary Knowledge Scale (Wesche & Paribakht, 1996). The results showed that the learners in both reading and listening conditions made noticeable lexical gains. In addition, the size of their vocabulary gains depended on their levels of text comprehension.

Summing up, the discussion above indicates that an adequate level of text comprehension plays two important roles in incidental L2 vocabulary acquisition from listening input. First, it helps to reduce the cognitive demands for the purpose of text comprehension and therefore learners will have more attentional capacity to process unfamiliar words in listening input. Second, it puts learners in a better position to make use of available contextual clues for their interpretation of these words. Therefore, if one would like to enhance the size of incidental L2 vocabulary gain from meaning-focused listening activities, one potential way is to facilitate L2 listening comprehension. This is also the main

¹ Test of English as a Foreign Language

way that I will use to inform my construction of pedagogical interventions in this research project. In the next section, I will first describe these interventions, and then justify how these interventions may foster L2 listening comprehension and therefore incidental vocabulary acquisition.

Intervention 1: Using Video rather than Audio Materials as Listening Input

The first pedagogical intervention that I suggest in this research project is to expose L2 learners to video rather than audio input. In fact, this factor will be used as a common feature across all studies in my research project. The effect of this factor on the size of incidental vocabulary uptake will be examined in Study 1. There are grounds for believing that a video text is more conducive to incidental vocabulary acquisition than an audio text.

First, previous research has provided evidence that video input can bring about better L2 listening comprehension than audio input. Baltova (1994), for example, compared L2 listening comprehension across three learning conditions: (a) viewing a video, (b) viewing the same video, but without the soundtrack, or (c) merely listening to the soundtrack. Text comprehension was measured by 16 multiple choice questions. The results showed that the learners in the two viewing groups (regardless of the presence of the soundtrack) obtained significantly better text comprehension than those in the listening-only group. Parry and Meredith (1984) also compared the effect of video and audio input on L2 listening comprehension. Text comprehension was determined by means of 60 multiple choice questions. The video group significantly outperformed the audio group. With a similar research design, Shin (1998) examined the extent to which video and audio input fostered L2 listening comprehension. The dependent measure included 31 open-ended content questions. The learners receiving video input again scored significantly higher on this measure than those receiving audio input. Wagner (2010) also investigated whether video input led to better L2 listening comprehension than audio input. Using a combination of 18 multiple choice and 22 short answer content questions as a dependent measure, he also found that the video condition indeed fared better. It is obvious from these studies that video input often results in better text comprehension than audio input.

There are two plausible explanations for this finding. First, video input offers a wide range of visual cues that are useful for schemata activation, but, by nature, unavailable in audio input. One important prompt is the images of the speakers, together with their non-verbal language patterns, such as: postures, gestures, and facial expressions. Non-verbal

language is often considered an integral part of the way an intended message is conveyed in spoken discourse. For instance, Sueyoshi and Hardison (2005) found that ESL learners at a North American university who were exposed to a videotaped lecture with visual access to the lecturer's non-verbal cues (i.e., gestures and face) outperformed those who were given the same lecture, but without access to non-verbal information on a text comprehension test. They concluded that video input should be more widely used in both instruction and assessment of L2 listening comprehension. Another visual prompt that can assist schemata activation is the physical setting in which the speech is given. Access to this information helps L2 learners to contextualize the input they are listening to in time, space, and to identify references speakers make to the surrounding context. Following Dual Coding Theory (Paivio, 1986), Vandergrift (2004) proposed the second explanation for the benefits of video input. According to this model, images can prompt L2 learners to access and activate relevant content schemata (i.e., top-down processing) faster than verbal language, which, in turn, gives them more time and attentional resources to devote to other aspects of listening such as parsing the speech stream (i.e., bottom-up processing). As argued above, this better text comprehension is, in turn, expected to assist L2 learners' interpretation of unfamiliar words.

Previous research has shown that video input is indeed a source of incidental L2 vocabulary uptake (e.g., Neuman & Koskinen, 1992; Rodgers, 2013; Sydorenko, 2010; Vidal, 2003, 2011). Neuman and Koskinen (1992), for example, found that L2 learners who watched an educational television program (regardless of the presence of L2 captions) made a larger incidental L2 vocabulary gain than those who simultaneously read and listened to the audio version of the same program or those who merely read its transcript. The availability of both verbal and pictorial clues in this TV program accounted for 21% of the variance in the size of vocabulary gains. The learners in Sydorenko's (2010) study indeed reported using pictorial clues in a given L2 audio-visual material in their efforts to interpret the meaning of unfamiliar words.

However, the vocabulary gain from this source is often found to be relatively small. Rodgers (2013), for instance, measured incidental L2 vocabulary uptake through watching a television series over a period of 13 weeks. Vocabulary gain was determined by means of two word-meaning recognition tests with different levels of difficulty. On both measures, the learners in the treatment condition were found to pick up around six out of 60 target words. Meanwhile, those in the control condition who did not view the series during the same period of time also learnt five target words. The effect size that the treatment condition in this study

created on vocabulary gain was relatively small with Cohen's d values of .49 on the difficult test and .43 on the easy test. Vidal (2011) compared the size of incidental L2 vocabulary gain from viewing three video-recorded lectures and reading their transcripts. Vocabulary gain was gauged by a modified version of Paribakht and Wesche's five-point Vocabulary Knowledge Scale (1993, 1997). The learners in the viewing condition obtained a mean gain of about six out of 36 target words in the immediate post-test. This gain dropped to three target words one month later. Those in the reading condition were found to make a significantly larger vocabulary gain of nine target words in the immediate post-test. This gain also considerably decreased to four words after one month. These studies show that L2 learners are able to develop their lexical knowledge through comprehension-based activities with video input. However, the vocabulary gain from this source still remains modest, especially as compared with the gain made from reading input (Vidal, 2011). In addition, there are learner-related factors (L2 proficiency and text comprehension), word-related factors (word type: technical, academic or low-frequency words), and input-related factors (quality of verbal and pictorial clues for word-meaning inference and frequency of word occurrences in input materials) that can influence the size of vocabulary gain. Thus, pedagogical interventions are also needed to enhance the size of lexical uptake from this source.

One potential intervention that previous research often recommends is to expose L2 learners to video input with textual support in the form of L1 and/or L2 captions. The presence of written texts in these materials is expected to help L2 learners to overcome their limitation in L2 lexical segmentation and recognition while dealing with listening input (van Zeeland, 2013). Thanks to this better L2 lexical segmentation and recognition, learners may improve their text comprehension (Brown et al., 2008; Webb & Chang, 2012). This better text comprehension is, in turn, likely to make them more able to process new lexical items in input materials (Webb & Chang, 2012, 2014). In addition, exposure to both spoken and written form of novel words in these materials is also believed to be more salient for learners' lexical form-meaning mappings than having only the spoken form as in the case of listening-only materials (Webb & Chang, 2012, 2014).

However, the question of whether captioned videos actually lead to a better vocabulary gain than uncaptioned videos has not been answered conclusively yet. Some studies in this area have reported a superior effect of captioned over uncaptioned videos on the size of lexical uptake (e.g., Huang & Eskey, 1999-2000; Koolstra & Beentjes, 1999;

Markham, 1999; Neuman & Koskinen, 1992). Other studies have contradicted this (e.g., Baltova, 1999; D'Ydewalle & Van de Poel, 1999; Karakas & Saricoban, 2012; Rodgers, 2013, Yuksel & Tanriverdi, 2009). To deal with such a conflicting finding, Montero Perez et al. (2013) carried out a meta-analysis with ten experiments (i.e., Baltova, 1999; Dunan, 1992; Hernandez, 2004; Huang & Eskey, 1999-2000; Hwang, 2003; Liversidge, 2000; Neuman & Koskinen, 1992; Sydorenko, 2010; Winke et al., 2010) to compare the degree to which captioned and uncaptioned videos fostered incidental L2 vocabulary uptake. The overall result supported captioned videos. They also found that the effect of captions on vocabulary gain was moderated by the type of vocabulary tests used in the aggregated experiments. In fact, this effect significantly differed in size when vocabulary gain was measured by a word recognition test or a word-meaning recall test. However, as Montero Perez et al. (2013) also acknowledged in their report, due to the small number of aggregated experiments in this meta-analysis, they did not examine the effects on the size of vocabulary gain of such other features of the vocabulary tests as the congruence in the presentation mode between the tests and the input materials – a methodological issue that has been raised in the second section of this chapter.

This review suggests that video input creates a better condition for L2 listening comprehension and therefore probably also for incidental vocabulary acquisition than audio input. However, the vocabulary gain from video input is still relatively limited, especially as compared to the gain from reading input. Therefore, pedagogical interventions that can enhance the size of lexical uptake from this source are still welcome. As mentioned, one potential intervention is to provide L2 learners with captioned videos. However, research investigating whether captioned videos bring about a larger vocabulary gain than uncaptioned videos to date have showed rather inconsistent results. In addition, many studies that claimed superiority of captioned over non-captioned videos for incidental L2 vocabulary acquisition used written word prompts in post-tests regardless of whether participants had watched non-captioned materials, which arguably introduces a confounding variable due to input-test (in)congruency (Sydorenko, 2010). In any case, teachers may consider the use of captions as temporary scaffolding in preparation for listening tasks unassisted by captions. Under that scenario, learners will sooner or later be required to cope with (and learn from) uncaptioned input. In the following section, I will suggest another intervention which is deemed to facilitate both L2 listening comprehension and incidental vocabulary uptake at the same time.

Intervention 2: Viewing a Video Twice rather than Once

The second intervention in this research project is to have L2 learners view the same video twice rather than once. Repeated exposure to the same listening input has long been considered a useful technique to make this input more comprehensible (e.g., Krashen, 1996; Vandergrift & Goh, 2012, pp.156-157). So far, research findings have supported this view.

Brown (2007), for example, invited 98 L2 learners to view five short videos twice. After each viewing session, these learners were required to rate their understanding of the video content from “nothing” (0%) to “everything” (100%). Their self-reported scores were found to significantly increase from the first to the second viewing. Inspired by Krashen’s idea of Narrow Listening (1996), Dupuy (1999) carried out an experiment with 255 L2 learners. In this experiment, the learners listened to short audio-recorded interviews with native speakers of French on familiar topics over one semester. They were allowed to repeat their listening to a particular topic as many times as they liked before moving on with another topic. At the end of the semester, they were asked to self-assess the usefulness of this practice for their L2 development. A majority of these learners reported that this activity helped them to improve not only their listening ability but their L2 vocabulary knowledge as well. Although these two studies clearly demonstrate the benefits of L2 repeated listening for text comprehension and vocabulary expansion, one obvious limitation of these studies is that they heavily relied on the learners’ self-evaluation of those benefits and did not triangulate these data with any objective dependent measure.

This methodological limitation was removed by a listening comprehension test in many other studies. Berne (1995), for instance, examined the effect of repeated listening on text comprehension. In this experiment, learners watched a short video-taped lecture twice and answered 10 multiple-choice comprehension questions after each viewing time. Their text comprehension scores after the second viewing were found to significantly surpass those after the first viewing. Cervantes and Gainer (1992) compared L2 listening comprehension across three conditions: (a) listening to a simplified version of a lecture once, (b) listening to the original version of this lecture once, and (c) listening to this original version, but doing it twice. Text comprehension was gauged by a gap-filling task. On this measure, the learners who listened to the simplified lecture once or the original lecture twice significantly outperformed those who listened to the original lecture, but only once. Using 10 multiple-choice questions as a dependent measure, Elkhafaifi (2005) gauged the degree to which L2

learners comprehended a video-recorded lecture after the first and second viewing. Their text comprehension scores significantly increased from the first to the second viewing. With a similar research design, Iimura (2007) also compared the levels of L2 listening comprehension after the first and the second exposure to the same input material. Text comprehension was determined by 20 open-ended questions. Again, there was a significant score improvement from the first to the second listening. These studies together reveal that exposure to the same listening input twice brought about significantly better text comprehension than exposure to it only once. However, these studies share another methodological limitation. They all used the same listening comprehension test twice. In this way, when it comes to the second listening, learners might direct their attention only to the information that they needed to answer the comprehension questions. Therefore, the difference in their test scores between the first and the second listening might not be due to their enhanced text comprehension, but due to the difference between a listening process with and without a clear listening purpose in mind.

This methodological issue was successfully dealt with in three studies by Chang and Read (2006), Lund (1991), and Saika (2009). Chang and Read (2006) removed this limitation by using a between-participant rather than a within-participant research design. Meanwhile, the other studies used an L1 written text-content recall task rather than a set of comprehension questions as a dependent measure. Chang and Read (2006) compared the effects of four listening support activities on L2 listening comprehension: (a) activating background knowledge for input topic before listening, (b) previewing comprehension questions before listening, (c) learning the form and meaning of unfamiliar lexical items before listening, or (d) listening to input materials twice. Two narrative texts were used as listening materials. Text comprehension was measured by 30 multiple choice questions. Input repetition was found to fare better than question and vocabulary preview, but not background activation. Lund (1991) compared the degree of L2 listening comprehension after the first and second listening, using a L1 text-content recall task as a dependent measure. The average number of recalled propositions from the input text was found to significantly increase from 2.9 after the first listening to 5.0 after the second listening. Saika (2009) also used a L1 text-content recall task as a dependent measure to compare the levels of L2 listening comprehension after the first and second listening. From the first to the second listening, the average number of idea units that learners in this study recalled also went up by 5.88 units for proficient learners and by 5.00 units for less proficient ones.

It is clear from these studies that repeated listening often cultivates significantly better text comprehension than one-time listening. As a way of catering for the limited capacity of human working memory (Cowan, 1995; Robinson, 2003; Schmidt, 1990), this type of listening gives L2 learners at least another listening time (which also entails more total ‘time on task’) to verify their previous interpretation of listening content and add information they have missed during their previous listening (Chang & Read, 2006; Lund, 1991). In addition, the announced task of exposure to listening input more than once helps learners to reduce their listening anxiety, which, in turn, makes them more mentally engaged with listening process (Brown, 2007). Therefore, having L2 learners view a video twice is also hypothesized to bring about better text comprehension than having them view the same video, but only once. With this better text comprehension, they are expected to stand a better chance of successfully interpreting the meaning of unfamiliar words, especially in their second viewing.

The second benefit of repeated listening for incidental vocabulary acquisition is its provision of the opportunity for L2 learners to reencounter novel words in input materials. According to Webb (2014), the more often unfamiliar words are encountered the more likely these words are acquired. Prior research has indicated a positive correlation between these two factors. Brown et al. (2008), for example, compared the size of incidental L2 vocabulary uptake among 28 target words that had four different bands of occurrences (i.e., 2-3, 7-9, 10-13, and 15-20 times) in three graded readers. The overall result showed that the more often the target words occurred in the input materials the more likely they were learnt. Horst et al. (1998) also examined this relationship with a group of 45 target words that occurred from 2 to 17 times in an audio graded reader. Vocabulary gain was indeed found to be positively correlated with word repetition: $r = .49$ (the p value was not reported). Rodgers (2013) also investigated this association. In his experiment, 60 unfamiliar words that occurred from 5 to 54 times in a television series were selected as the target words. Vocabulary gain was measured by two versions of a word-meaning recognition test – one version was tougher than the other (see above). On the tough measure, the vocabulary gain was positively correlated with the frequency of word occurrences in the listening input: $r = .30$ ($p < .05$). On the easy measure, this relationship was also positive, but fell short of the required significance level: $r = .18$ ($p = .16$). Van Zeeland and Schmitt (2013) compared the size of lexical uptake across 24 words that had different frequency of occurrences of 3, 7, and 11 or 15 times in four listening passages. Again, word repetition had a significant effect on the size of vocabulary

gain. Vidal (2003, 2011) was also interested in this association. In these studies, 36 unfamiliar words that occurred from one to six times across three video-recorded lectures were chosen as the target words. Lexical gain was measured by a self-report test using a modified version of Paribakht and Wesche's Vocabulary Knowledge Scale (1993, 1997). Word repetition was found to explain 24% of the variance in the vocabulary gain scores.

These studies consistently show a positive correlation between the frequency of word encounters and the size of vocabulary gain from listening input. A difference with previous work on such frequency effects is that, in the case of repeated listening to the same text, the words are re-encountered in identical contexts, which logically entails fewer opportunities for learners to fine-tune their word-meaning inference thanks to an accumulation of contextual clues (R. Ellis, 1999). However, on the plus side, this type of word reencounters allows learners to make use of their previous understanding of input content for their lexical form-meaning mappings.

Together, having L2 learners view a video twice rather than once can provide two distinct benefits for incidental L2 vocabulary acquisition. First, it enhances text comprehension. With this better text comprehension, learners are more able to infer the meaning of unfamiliar words out of context, especially in their second viewing. Second, this viewing procedure also allows them to reencounter these words in the same contexts, which helps them to deploy their previous text comprehension for word-meaning inferencing. The combined effect of these two factors on vocabulary gain will be measured in Study 2 in this research project.

Intervention 3: Adding an Output Task between these Two Viewing Times

In this intervention, L2 learners will be required to view a video, make an oral summary of the content, and then view this video again. There are both theoretical and empirical reasons to predict that this oral summary activity will help learners to enhance their text comprehension and vocabulary gain at the same time.

The integration of the summary activity between two times of viewing the same video is theoretically based on Wittrock's Model of Generative Teaching of Comprehension (1991). According to this model, text comprehension is an active process in which learners construct given meaning through *generating* mental links (a) between different ideas within a given input text and (b) between these ideas and relevant schemata in their long-term memory.

Therefore, text comprehension can be much facilitated through classroom activities that directly induce this generation (Gambrell, Pfeiffer, & Wilson, 1985). One activity that previous research often recommends is retelling input content in the form of a written or oral summary activity (Linden & Wittrock, 1981). This activity is deemed to prompt learners to relate information from one section of an input text to another (integration of information) and also to their prior knowledge and experience (personalization of information) (Gambrell, Pfeiffer, & Wilson, 1985; Wittrock, 1981). Therefore, it can work as a generative tool to foster text comprehension. According to Wittrock (1991), metacognition is an external, but facilitative factor that can help to enhance the effect of this tool on learners' text comprehension.

In fact, previous research has provided supporting evidence for the usefulness of using a summary activity as a generative tool to facilitate text comprehension. Doctorow, Wittrock and Marks (1978), for example, carried out a large-scale classroom experiment with 500 L1 child learners. This experiment used a between-participant research design to compare the effects of three reading conditions on text comprehension. In two treatment conditions, the learners were required to read a set of short texts and then generate a summary sentence for each paragraph in those texts by using either the information or the heading given in that paragraph. Those in the control condition read the same texts, but were not asked to generate any summary sentences. The time on task was kept constant across all conditions. The results showed that both treatment conditions led to significantly better comprehension and retention of the reading content than the control condition. Gambrell, Koskinen, and Kapinus (1991) also examined the effect of summarizing input content after reading on text comprehension. In this study, 48 L1 children learners were involved in four reading sessions over a period of two weeks. In each session, these learners were asked to read one narrative story and then summarize its content. Their text comprehension was measured by the number of propositions and structure elements that they could recall from the story and their responses to eight content questions. The difficulty level of the stories was kept roughly the same across all reading sessions. The results showed that, from the first to the fourth reading session, these learners significantly improved their text comprehension on all measures. Kelly and Wittrock (1984) compared text comprehension across three reading conditions: (a) reading short texts and then summarizing reading content, (b) reading the same texts and then drawing the relationship between different propositions in reading

content, and (c) reading the texts only. On the final comprehension test, both summarizing and drawing groups were found to significantly outperform the reading-only group.

These three studies clearly demonstrate the benefit of using a summary activity as a generative tool to foster text comprehension. It should be, however, noted that these studies have all examined this effect in the context of L1 reading, but not in the context of L2 listening. Thus, it may be interesting to examine whether this benefit also holds true in the case of L2 listening. In the studies reviewed above, learners were not given another exposure to the input materials to self-assess the quality of their text comprehension. Put differently, they had fewer opportunities to further develop their metacognitive strategies – a facilitative factor of generative learning for text comprehension that Wittrock (1991) suggested in his model. Another limitation of these studies is that they are all quantitative in nature. Due to this limitation, these studies have not provided any direct evidence for the cognitive and metacognitive processes that learners used to deal with the summary task and how these processes affected their text comprehension. Therefore, more empirical evidence on the effect of using a summary activity as a generative tool to foster text comprehension, especially in the case of L2 listening comprehension, is still required. With the present intervention, I would like to remove the first two limitations mentioned above. The last limitation concerning the direct evidence for the cognitive and metacognitive processes that learners use to cope with the summary task and how these processes affect their text comprehension and incidental vocabulary uptake will be dealt with in the final study of this research project.

As an output task, this summary activity can stimulate incidental vocabulary acquisition in accordance with the three main tenets of Swain's (1985, 1995, 2005) Output Hypothesis. First, it promotes "*noticing*" (Swain, 1995, p.125). When learners are required to summarize the video content, they may notice gaps in their lexical resources and subsequently direct their attention to relevant lexis as they view this video again. Second, this output task creates the opportunity for them to experiment with using a newly met word in a particular context and therefore helps them to consolidate their knowledge of this lexical item – a benefit that Swain refers to as the *hypothesis-testing* function of output tasks. The third tenet of this hypothesis, which has perhaps received less attention in the research literature, is that output *enhances fluency* of L2 use. That is because using a recently encountered word in an output activity fosters memorization of this lexical item, which, in turn, increases the degree of automaticity in recognizing it from a given input text and also retrieving it from memory for communicating ideas.

Previous research has furnished evidence for the benefit of output tasks for incidental L2 vocabulary acquisition. De la Fuente (2002), for example, examined the size of incidental L2 vocabulary gain from two listen-and-do tasks. In these tasks, learners listened to a native speaker of the target language giving directions to put objects (which were also the target words) in corresponding places in a picture. In the pre-modified input condition, learners were not allowed to interact with the native speaker, but they were given the meaning of the target words right in the directions. In the negotiated input condition, learners were not given the meaning of the target words in the directions, but they were allowed to ask the native speaker for the meaning of these lexical items if they liked. In the negotiated input plus pushed output condition, the procedure was the same as that in the negotiated input condition. However, after the first listening, learners took the role of the direction providers. The time invested was kept the same at 40 minutes across all conditions. Vocabulary gain was measured by Wesche and Paribakht's Vocabulary Knowledge Scale (1997). The negotiated input plus pushed output group was found to obtain not only a better text comprehension but also a larger vocabulary gain than the other groups.

R. Ellis and He (1999) also measured the size of incidental L2 vocabulary uptake from a similar listen-and-do task. In this task, learners listened to either their teacher or their friend giving directions to put pieces of furniture (which were also the target words) in their corresponding places in a picture under three learning conditions. The learners in the pre-modified input condition listened to their teacher giving the directions. They were given the meaning of the target words in the directions, but they were not allowed to interact with their teacher. The learners in the interactionally modified input condition were not given the meaning of the target words in the directions, but they were allowed to ask their teacher for the meaning of these lexical items if they needed. The learners in the interactionally modified output condition listened to their teacher explaining the meaning of the target words before they worked in pairs and took turns to give their own directions to their friends. In this condition, their friends were also allowed to ask for the meaning of the target words if they needed. The time on task was kept the same at 45 minutes across all conditions. Vocabulary gain was determined by means of a receptive test and a productive test. Again, the interactionally modified output condition brought about a significantly better text comprehension and larger vocabulary gain than the other conditions. These studies both show that using an output task (either with or without interaction support) after a meaning-focused

listening activity led to better text comprehension and incidental vocabulary acquisition than the effect of the meaning-focused listening activity alone.

However, it should be also noted here that the nature of the output tasks in the studies reviewed above and that of the output task in my proposed intervention are distinctively different, which may lead to differing effects on the size of vocabulary gain. Let me start with the output task in the study by R. Ellis and He (1999). In this task, learners were first explicitly instructed on the form and the meaning of target words that labelled some objects and then required to use these words to tell their friends where to put these objects in their corresponding places in a given picture. The focal attention of the learners in this case was directly placed on the target words. Thus, this output task is arguably what Long (1988, 1991) and Long and Crookes (1992) referred to as a “*focus on formS*” task. In the output task that de la Fuente (2002) used in her study, learners used the words (which also labelled real objects) that they themselves had picked up from their preceding bidirectional listening task to tell their interlocutor where to put these objects in their corresponding places in a provided picture. During both listening and speaking tasks, the learners in this study mainly focused on the input and output content, but their attention was occasionally diverted to the form and/or the meaning of the target words when communication problems caused by these lexical items arose. Therefore, with reference to Long (1988, 1991), Long and Crookes (1992), and Williams (2005), such a task can be considered a “*focus on form*” task. However, as the learners were told beforehand to use the target words during the output task and this task was specially designed in a way that communication problems (if any) were likely caused by these words only, this output task has a “*planned*” (Doughty & Williams, 1998; Harley, 1998; Swain, 1998) rather than an “*incidental*” (R. Ellis, 2001) *focus on form*. The output task in my proposed intervention requires learners to summarize input content using whatever lexical resources they have. In case they recognize that they need some words from the listening input to convey their ideas and then possibly direct their attention to these words in their second listening, this engagement with vocabulary is, by nature, triggered by their own communication needs and problems. In other words, this lexical uptake is purely a result of *incidental focus on form* (R. Ellis, 2001). Given the importance of attention and/or awareness in language acquisition (Robinson, 2003; Schmidt, 2001), the effect size of my output task on vocabulary gain might be smaller than those of the output tasks in previous research. However, one cannot deny the potential that this task has for incidental vocabulary uptake,

especially the “*noticing*” and “*fluency enhancement*” affordances that Swain (1985, 1995, 2005) suggested in her Output Hypothesis.

In my proposed intervention, learners are asked to summarize the video content that they have just viewed. This output task also gives them the opportunity to experiment with using words that they encounter for the very first time during their first viewing. They may use these words either verbatim or in new syntagmatic contexts that differ from the original phrasings in the given video. In both cases, such use is believed to foster their memorization of these lexical items (Swain, 2005). However, as already mentioned, a core tenet in Wittrock’s Model of Generative Learning and Teaching (1991) is that learners will understand and remember new information (e.g., novel words) better if they attempt to generate mental connections between this new information to the relevant schemata in their long-term memory. This may tie in with Nation’s notion of Vocabulary Generation (2013), according to which learning tasks that give learners opportunities to encounter (i.e., receptive generation) or use (i.e., productive generation) recently learnt words in new contexts benefit vocabulary retention more than tasks in which they meet or produce the words in the same contexts. This view has been supported by the results from Joe’s study (1998). In this study, Joe compared the size of incidental L2 vocabulary uptake from a reading and retelling task across three learning conditions: (a) reading a text with explicit training on how to use background knowledge to interpret the meaning of new words or new content presented in the text and then retelling the text content without access to the original text, (b) reading the same text without such training, but with access to this text when retelling the content, and (c) having neither such training, reading nor retelling (control condition). Vocabulary gain was measured by a self-report test using a modified version of Wesche and Paribakht’s Vocabulary Knowledge Scale (1993) and two versions of a word-meaning recognition test (one was easier than the other). Both treatment conditions were found to yield a significantly better vocabulary gain than the control condition. The target words that were actually used by the learners in their retelling session were retained significantly better in the post-tests than the words that were not used at all. The words used in contexts that linguistically differed from their original contexts in the reading materials were also remembered significantly better than the ones that were repeated verbatim.

As learners in my proposed intervention are given another viewing of the same video right after their summary activity, this intervention also creates a favourable condition for incidental vocabulary acquisition in accordance with Laufer and Hulstijn’s Involvement Load

Hypothesis (2001). In this hypothesis, the more deeply a learning task prompts L2 learners to process new lexical items the better they can learn and remember these words. The depth of lexical processing that a task induces is operationalized by the strength of three task components: need, search and evaluation. For example, need is considered moderate when the teacher asks learners to find the meaning of an unfamiliar word (i.e., extrinsic motivation). However, need is strong when learners themselves want to find the word meaning to solve their own comprehension problem (i.e., intrinsic motivation). Search is also rated as moderate or strong, but depending on whether learners are required to retrieve a word meaning from a given word form or a word form from a given word meaning, respectively. As for evaluation, if learners are asked to make a comparison between different shades of meaning of a new word, this evaluation is regarded as moderate in terms of strength. Evaluation is strong if they need to decide whether a specific meaning fits a specific context. In the case of my proposed intervention, as mentioned above, when learners are summarizing listening content, they may need some words that they have encountered during their first viewing, but have not picked up yet. This need can, in turn, prompt them to search for these lexical items during their second viewing. If these items are indeed detected in the second viewing, learners also need to evaluate whether these items can convey their intended meaning in their summaries. Following Laufer and Hulstijn's Involvement Load Hypothesis (2001), need, search and evaluation are all strong in this case. In other words, my proposed intervention is very likely to induce high involvement load in terms of lexical processing and therefore to foster incidental vocabulary uptake.

Previous research has found that the higher the involvement load a learning task induces the stronger the effect it has on vocabulary acquisition and retention. Hulstijn and Laufer (2001), for instance, compared the effects on vocabulary gain of three learning tasks with different involvement loads: (a) reading a short text for overall comprehension, (b) reading the same text for overall comprehension, but followed by a gap-filling activity eliciting target words, and (c) writing a paper using target words. According to the Involvement Load Hypothesis, the depth of lexical processing in these tasks increased from the reading task to the reading plus gap-filling task and then to the writing task. As expected, vocabulary gain was found to be the largest in the writing task, smaller in the reading plus gap-filling task, and the smallest in the case of the reading only task. One limitation of this study is that it did not control the time on task. Thus, the differing effects of these three tasks may be also attributed to the different amount of time that learners processed the given target

words in each task (Folse, 2006; Keating, 2008). Kim (2008) replicated the study by Hulstijn and Laufer (2001), but kept the time invested the same at 40 minutes across the three learning tasks. The results from this experiment were consistent with what Hulstijn and Laufer (2001) found. In another experiment, Kim (2008) compared the effects on lexical uptake of two learning tasks which were deemed to have a similar involvement load: (a) writing an essay using target words and (b) writing sentences using target words. These tasks indeed brought about a similar amount of vocabulary acquisition and retention.

The studies reviewed above suggest that a learning task which induces high involvement load in terms of lexical processing may foster both vocabulary acquisition and retention. As my proposed intervention is believed to provide learners with strong need, search and evaluation regarding relevant content words, this intervention is also expected to produce a positive effect on learners' vocabulary gain. However, as discussed above, my proposed intervention can only prompt *incidental* rather than *planned focus on form* as in the case of the studies by Hulstijn and Laufer (2001) and Kim (2008). In addition, the real-time nature of listening (Buck, 2001) does not allow learners in my intervention to spend extensive time processing a particular lexical item at the input stage. Thus, the more successfully my learners understand the video content right from their first viewing the more mental resources they will have in the second viewing to process words the need for which was experienced during the preceding summary activity.

In this field, there have been several studies that directly examined the effect of using a summary activity as a text-based output task on incidental L2 vocabulary uptake. Lu (2013), for example, carried out a classroom experiment with 122 L2 learners. These learners were required to read a passage in which nine unfamiliar words were highlighted and their meaning was glossed at the text margin. After the reading session, the learners were asked to use these nine words either to (a) fill the blanks in a set of nine sentences, (b) fill the blanks in three sets of nine sentences, (c) fill the blanks in a summary of the reading text, or (d) write a summary of the reading text. Vocabulary gain was gauged by a word-meaning translation test, a word spelling test, and a word use test. The learners in the summary condition were found to make a significant vocabulary gain on all measures. This gain was not significantly larger than those in the other conditions in the immediate post-test, but in the delayed post-test. Rassaei (2015) also compared the effects of three reading-based output activities on incidental L2 vocabulary acquisition. In this study, learners were required to read two passages, in each of which seven unfamiliar words were highlighted and their meaning was

glossed in L1. After the reading session, they were asked to use these words either to (a) write a summary of the text content, (b) generate content questions and then answer them, and (c) write a prediction of how the text content would continue. Lexical uptake was gauged by two gap-fill tests. The learners in the summary condition were found to obtain a significant vocabulary gain in the post-test. However, this gain did not significantly surpass those in the other conditions.

Taken together, the integration of the oral summary activity between two times of viewing the same video is first expected to give L2 learners the opportunity to reprocess their perceived information from the first viewing and then get prepared for their second viewing, which, in turn, may help them to increase their text comprehension. With this better text comprehension, learners are expected to be more able to process unfamiliar words in the video, especially in the second viewing. As a text-based output task, this summary activity creates a favourable condition for learners to use newly met words and therefore consolidate their knowledge of these lexical items. It may also help them to notice gaps in their lexical resources and then prompt them to direct their attention to relevant lexis in the second viewing. The effect of this intervention on the size of vocabulary gain will be measured in Study 3 in this research project.

Intervention 4: Viewing Three Videos on Similar Topics

In this intervention, L2 learners will be asked to view a set of three videos on similar topics - a viewing procedure that will be hereafter called, for short, “*Narrow Viewing*” as an analogy to Krashen’s Narrow Reading (2004) and Narrow Listening (1996). This Narrow Viewing is expected to foster L2 listening comprehension in two different ways.

First, it is expected to enhance learners’ text comprehension thanks to their increasing familiarity with the given topics. Following Schemata Theory (Rumelhart, 1980), when learners are familiar with listening topics they are more able to make conscious use of relevant schemata available in their long-term memory to facilitate their listening process. The claim is that by drawing on relevant schematic knowledge, learners can make more accurate inferences about listening content that they have not been able to fully process due to gaps in language knowledge and/or speech processing capacity. Given the limited capacity of human working memory and real-time nature of listening, this process of bringing meaning to the text through schema activation is claimed to play a critical role in L2 listening comprehension. Relevant schemata allow learners to map meaning directly to long-term

memory thus reducing demands on their attentional capacity. This capacity is then freed up for bottom-up processing, for noticing learnable features in the input and/or for conscious attention to cognitive and metacognitive strategy use (Vandergrift, 2007).

Previous research has shown that topic familiarity has a positive impact on L2 listening comprehension. Leaser (1998), for example, compared the degree of text comprehension from listening to two passages with different levels of topic familiarity. Text comprehension was measured by a L1 written text-content recall test and a multiple-choice comprehension test. The effect of topic familiarity on L2 listening comprehension was significant in the case of the L1 written text-content recall measure, but not in the case of the multiple-choice comprehension measure. Long (1990) also examined the effect of topic familiarity on L2 listening comprehension. In this study, text comprehension was determined by means of a L1 written text-content recall test and a True/False comprehension test. Again, the effect of topic familiarity was found to be significant on the first, but not on the second measure. Sadighi and Zare (2006) also gauged the effect of topic familiarity on L2 listening comprehension. Learners were required to listen to five TOEFL listening passages. Half of the learners were well-prepared for input topics before listening while the other half were not. Text comprehension was measured by 50 typical TOEFL listening comprehension questions. The results showed that topic preparation indeed produced a positive effect on L2 listening comprehension. Schmidt-Rinehart (1994) also compared the level of text comprehension made from listening to a familiar topic and an unfamiliar topic. Text comprehension was gauged by a L1 written text-content recall test. Listening to the familiar topic again yielded significantly better text comprehension than listening to the unfamiliar topic. Zidong (1998) also investigated the effect of topic familiarity on L2 listening comprehension, but together with that of L2 proficiency level and type of comprehension question. Only topic familiarity and L2 proficiency level were found to have a significant effect on text comprehension, but the effect of topic familiarity was far larger than that of L2 proficiency. The five studies above consistently show a positive effect of topic familiarity on L2 listening comprehension. Having L2 learners watch three TED Talks videos on similar topics is therefore expected to enhance their text comprehension thanks to their increasing familiarity with the topics when they move on from one video to another.

In addition, exposure to several input texts on similar topics may also help learners to increase their lexical coverage or proportion of known words. Previous research has provided evidence for this in both reading and listening contexts. Kyongho and Nation (1989), for

example, compared the lexical profiles between two collections of topic-related and topic-unrelated news, which shared a similar number of running words. The results showed that if learners had a vocabulary size of 2,000 most frequent word families in L2 at the outset of their reading, they would encounter much fewer unfamiliar word families in the former than in the latter. Reading topic-related news also significantly reduced the number of unfamiliar word families that they encountered from the first to the final text. Meanwhile, the number of such unfamiliar word families remained roughly the same across all texts in the other collection. Rodgers and Webb (2011) examined the degree to which word families reoccurred in theme-related and theme-unrelated TV programs. Using the RANGE program (Nation and Heatley, 2002), they compared the lexical profile of 142 episodes taken from a single season of six individual programs (theme-related) with that of 146 other episodes taken from six sets of randomly selected programs (theme-unrelated). The theme-related TV programs were found to contain much fewer word families than the theme-unrelated ones. If learners already had a vocabulary size of 3,000 most frequent word families in L2, they would encounter a noticeably smaller number of unfamiliar word families in the theme-related programs than in the theme-unrelated programs. Schmitt and Carter (2000) also compared the lexical profiles between two sets of topic-related and topic-unrelated stories in newspapers. Again, L2 learners would encounter fewer word types in the former than in the later. Based on the premise that high-frequency content words (that were likely to be already known by L2 learners) were more important for text comprehension than functional words (e.g., articles like “a”, “an”, and “the” and prepositions like “in”, “on”, and “at”) and low-frequency content words, Schmitt and Carter compared the number of high-frequency content words between the two collections. The result actually showed that more high-frequency content words reoccurred in the topic-related stories than in the topic-unrelated stories.

These three studies clearly indicate that exposure to several input texts on similar topics reduces the number of potentially unfamiliar word families that L2 learners encounter or, in other words, increases their lexical coverage of those texts. Previous research has found a positive association between lexical coverage and text comprehension in both reading (Hsueh-Chao & Nation, 2000; Kurnia, 2003; Laufer, 1989; Liu & Nation, 2000) and listening (Bonk, 2000; Stæhr, 2009) context. Therefore, a *Narrow Viewing* procedure is also expected to facilitate L2 listening comprehension in this regard.

Indirect support for this comes from a study by Pulido (2007), who measured the effects of topic familiarity and lexical coverage on incidental L2 vocabulary gain through

reading. In this study, 35 L2 learners of Spanish were required to read two narrative passages with different degrees of topic familiarity and lexical coverage. Vocabulary gain was gauged by a word meaning recall test and a word meaning recognition test. Both topic familiarity and lexical coverage were found to have significant effects on the size of lexical uptake and retention. Therefore, it might be worth examining whether these two factors also influence incidental L2 vocabulary gain from listening.

To sum up, in this intervention, L2 learners will view a set of three videos on similar topics. This *Narrow Viewing* procedure is deemed to enhance learners' text comprehension thanks to their increasing familiarity with the given topics and their increasing lexical coverage of the input materials when they move on from one video to another. With this better text comprehension, they will have more mental resources available to process new lexical items in the given input materials (Boers et al., 2007) and more contextual clues for their word-meaning interpretation (Haastrup, 1991; Nation, 2001, 2013; Stæhr, 2009). The effect of this intervention on the size of incidental vocabulary gain will be measured in Study 4 in this research project.

Intervention 5: Adding an Output Task into this Narrow Viewing Procedure

The last intervention in this research project is to have L2 learners view either the first or second half of the three-video set in the previous intervention, take turn to share the content with a classmate who has not watched the same half yet in an interactional summary activity, and then move on with the other half. This intervention is expected to facilitate L2 listening comprehension and incidental vocabulary acquisition in several ways.

First, like the *Narrow Viewing* intervention (see the section above), the present procedure is also expected to help L2 learners to enhance their text comprehension thanks to their increasing familiarity with the given topics (Leeser, 1998; Long, 1990; Sadighi & Zare, 2006; Schmidt-Rinehart, 1994; Zidong, 1998) and increasing lexical coverage of the input materials when they move from one video to another (Kyongho & Nation, 1989; Rodgers & Webb, 2011; Schmitt & Carter, 2000).

Following Wittrock's Model of Generative Teaching of Comprehension (1991), the summary activity in this intervention is also likely to prompt learners to generate mental links between different ideas within the listening content and between this content and their

previous knowledge and experience, which may, in turn, enhance their comprehension and retention of the listening content.

One special feature in this summary activity is that learners will receive a summary of the video content from their classmate before they are exposed to the input materials themselves. In this way, learners can form an “advance organizer” of upcoming listening content, which is believed to facilitate their actual listening comprehension (Ausubel, 1960, 1978). Previous research has consistently supported this view. Herron (1994), for example, examined the effect of providing a summary of input content as an advance organizer on L2 listening comprehension. In this study, learners watched a series of videos either with or without such advance organizers. Text comprehension was indeed significantly better when learners received such a summary before viewing. Herron, Cole, York, and Linden (1995) compared L2 listening comprehension across three groups of learners. Two groups received either a summary of the video or multiple-choice questions about its content before watching it, while the third group received no such advance organizers. The former two groups subsequently scored significantly higher on a text comprehension test than the latter. Jafari and Hashim (2012) also compared L2 listening comprehension across three learning conditions. In this study, learners were required to listen to short passages, but receiving either (a) a summary of the input content, (b) a set of key words in the input materials or (c) no support before listening. The results showed that the learners who received the key vocabulary or the summary of the input content before listening significantly outperformed those who did not receive any pre-listening support. The effects of the summary and the key vocabulary were roughly the same. Follow-up interviews with the learners, however, revealed that they preferred receiving the summary to the key words. It was because the summary helped them to grasp the topic and main ideas of the upcoming listening content, which, in turn, facilitated their input processing. Meanwhile, they considered the key words useless and even distractive to their listening process. These studies together show that receiving a summary of input content before listening has great potential for fostering text comprehension.

It should be also noted that while learners in my proposed intervention are listening to their classmate summarizing the content of the upcoming video(s) they have the opportunity to negotiate this content with the summary giver. Following Long’s Interaction Hypothesis (1980, 1983, 1985), such a negotiation is likely to render input more comprehensible and therefore create a favourable condition for second language acquisition, including incidental

vocabulary uptake. In his latest version of Interaction Hypothesis, Long (1996) added that, in such a negotiation episode, learners may negotiate not only input content but also new language codes (e.g., the form and/or meaning of a novel word), which can also result in their uptake of those codes as well. So far, research has strongly supported this hypothesis.

R. Ellis et al. (1994), for example, carried out two experiments to measure incidental L2 vocabulary gain from a listen-and-do task. In this task, learners listened to a native speaker of the target language giving directions to put kitchenware items (which were also the target words) in their correct places in a kitchen map in three learning conditions. In the baseline input condition, the directions were not modified (that is why this condition was referred to as the baseline input condition). In the pre-modified input condition, the native speaker pre-modified the directions by adding the meaning of the target words and then read these directions aloud. In the interactionally modified input condition, learners listened to the unmodified directions, but they could halt the native speaker to ask for the meaning of the target words when needed. Both pre-modified and interactionally modified input groups were found to obtain a significantly better text comprehension and larger vocabulary gain than the baseline input group. The interactionally modified input group also gained better scores on the text comprehension and vocabulary post-tests than those in the pre-modified input group, but the difference fell short of the required significance level in the case of the vocabulary post-tests. In the second experiment, another group of L2 learners undertook exactly the same experimental procedure as described above. However, for some unknown reason, the data related to the baseline input condition could not be retained. The results showed that the interactionally modified input condition again yielded a significantly better text comprehension and larger vocabulary gain than the pre-modified input condition.

Using a similar research design, Loschky (1994) also compared the size of incidental L2 vocabulary uptake across three listening conditions: (a) baseline input, (b) pre-modified input, and (c) interactionally modified input. Vocabulary gain was measured by a Yes/No word-form recognition test. All conditions were found to bring about significant vocabulary gains. These gains were, however, not different in size. This might be attributed to the lack of discriminatory power of the vocabulary measure in this study. Obviously, the main aim of the pre-modified input and the interactionally modified input was to help the learners in this experiment to establish the form-meaning connection for new words. However, the vocabulary test merely focused on whether they could recognize the word form. Testing their knowledge of word meaning or, to be more precise, form-meaning connection rather than

merely testing recognition of word form might therefore have been more appropriate for detecting differential effects of the three learning conditions.

Yi and Sun (2013) also investigated incidental L2 vocabulary uptake from pre-modified and interactionally modified listening input. Two experiments were carried out with two different groups of L2 learners. Vocabulary gain was measured by a word-meaning recall test before the experiment and a gap-fill test after the experiment. In both experiments, the interactionally modified input group again significantly outperformed the pre-modified input group on the vocabulary post-test. Vocabulary gain in this study was also found to be positively correlated with text comprehension.

It is clear from these studies that when learners are allowed to interact with their interlocutors they can negotiate input content and new vocabulary knowledge, which, in turn, both foster their incidental vocabulary uptake. Thus, my proposed interactional summary activity is also expected to create such a favourable condition for vocabulary gain.

Like in the case of Intervention 3, the proposed summary activity in the present intervention may also stimulate incidental vocabulary acquisition in accordance with Hulstijn and Laufer's Involvement Load Hypothesis (2001), Nation's Vocabulary Generation (2013), and three functions of Swain's Output Hypothesis (1985, 1995, 2005): noticing, hypothesis-testing, and fluency development (please see Intervention 3 above for a full discussion). However, unlike the unidirectional summary activity in Intervention 3, interaction is encouraged in the present summary activity. Thus, learners can directly negotiate the form and meaning of unfamiliar words with their classmates. This negotiation is therefore expected to prompt them to reflect on their lexical knowledge, identify problems when this knowledge is put to use, and then seek feasible solutions to these problems. Through this process of reflection and problem-solving, learners might build up further knowledge "*about language*" (Swain, 2005, p.478). This is also the final function in Swain's Output Hypothesis – *Metalinguistic function*.

In summary, this intervention is first expected to foster learners' text comprehension thanks to their increasing familiarity with the listening topics and increasing lexical coverage of the listening input as they move on from one video to another. The interactional summary activity in this intervention is deemed to further enhance their text comprehension in two different ways. On the part of the summary giver, they have the opportunity to reprocess information retained from their previous listening and therefore improve their comprehension

and memorization of the listening content (Wittrock, 1991). On the part of the summary receiver, they are given a summary of input content that can serve as an advance organiser before listening (Ausubel, 1960, 1978). They are also allowed to negotiate with the summary giver about both the form (new language codes) and the meaning (content) of the summary. This negotiation is believed to facilitate not only their text comprehension but also their incidental vocabulary uptake (Long, 1996). As an interactional output task, this summary also creates a favourable condition for learners to pick up new lexical knowledge in line with Hulstijn and Laufer's Involvement Load Hypothesis (2001), Nation's Vocabulary Generation (2013), and Swain's Output Hypothesis (1985, 1995, 2005). Study 5 in this research project will gauge the effect of this intervention on the size of incidental L2 vocabulary gain.

Summary

The review above shows that incidental L2 vocabulary acquisition from listening-only input is limited. Therefore, pedagogical interventions are needed to enhance the size of vocabulary gain from this source. One way to do so is to develop pedagogical interventions that foster text comprehension. In this way, learners will have more mental resources for their interpretation of unfamiliar words in input materials. In this research project, I have suggested five interventions, which are deemed to facilitate text comprehension in different ways. There has been both theoretical and empirical ground to support these suggested interventions. This ground is summarized in Table 2 below:

Table 2: Suggested interventions in this research project		
Intervention	Ground	
Having video rather than audio materials as listening input	Text comprehension	Presence of pictorial support in video input facilitates text comprehension.
	Theoretical ground	Dual coding of information (Jones & Jan, 2002; Sadoski, 2005)
	Empirical ground	Baltova (1994); Parry & Meredith (1987); Sueyoshi & Hardison (2005); Shin (1998); Wagner (2010)
	Vocabulary uptake	Pictorial clues in video input also foster lexical form-meaning mappings.
	Theoretical ground	Dual coding of information (Jones & Jan, 2002; Sadoski, 2005)
	Empirical ground	Neuman & Koskinen (1992); Sydorenko (2010)
Viewing	Text comprehension	Another viewing time makes given input more comprehensible.

	Theoretical ground	Repeated listening (Krashen, 1996)
	Empirical ground	Bern (1995); Brown (2007); Cervantes & Gainer (1992); Chang & Read (2006); Dupuy (1999); Lund (1991); Saika (2009)
	Vocabulary uptake	Word reencounters facilitate incidental vocabulary uptake.
	Theoretical ground	Word repetition in incidental vocabulary acquisition (Webb, 2014)
	Empirical ground	Brown et al. (2008); Horst et al. (1998); Rodgers (2013); van Zeeland & Schmitt (2013); Vidal (2003, 2011)
Adding a summary activity between the two viewing times	Text comprehension	A summary activity can be used as a generative tool to foster text comprehension.
	Theoretical ground	Wittrock's Model of Generative Teaching of Comprehension (1991)
	Empirical ground	Doctorow et al. (1978); Gambrell et al. (1991); Kelly & Wittrock (1984)
	Vocabulary uptake	Incidental vocabulary acquisition can be stimulated by such lexical processing in output tasks as: noticing, fluency enhancing, hypothesis testing, and generative processing.
	Theoretical ground	Hulstijn & Laufer's Involvement Load Theory (2001); Nation's Vocabulary Generation (2013); Swain's Output Hypothesis (1985, 1995, 2005)
	Empirical ground	de la Fuente (2002); R. Ellis & He (1999); Hulstijn & Laufer (2001); Joe (1998); Kim (2008); Lu (2013); Rassaei (2015)
Viewing three videos on similar topics	Text comprehension	This intervention may enhance learners' text comprehension thanks to their increasing topic familiarity and lexical coverage when they move on from one video to another.
	Theoretical ground	Krashen's Narrow Listening (1996)
	Empirical ground	Bonk (2000); Leaser (1998); Long (1990); Rodgers & Webb (2011); Sadighi & Zare (2006); Schmidt-Rinehart (1994); Schmitt & Carter (2000); Stæhr (2009); Zidong (1998)
Adding an interactional summary activity into the procedure	Text comprehension	This intervention may help learners to improve text comprehension thanks to: (a) their increasing topic familiarity, (b) their increasing lexical coverage, (c) their reprocessing of perceived information from the previous listening, (d) their listening to a summary of input content as an advance organizer before listening, and (e) their negotiation with their interlocutor about the summary

		content.
	Theoretical ground	Krashen's Narrow Listening (1996); Wittrock's Model of Generative Teaching of Comprehension (1991); Ausubel's Advance Organizer (1960, 1978), Long's Interaction Hypothesis (1980, 1983, 1985, 1996)
	Empirical ground	Bonk (2000); Herron (1994); Herron et al. (1995); Jafari & Hashim (2012); Leeser (1998); Long (1990); Rodgers & Webb (2011); Sadighi & Zare (2006); Schmidt-Rinehart (1994); Schmitt & Carter (2000); Stæhr (2009); Zidong (1998)
	Vocabulary uptake	Incidental vocabulary acquisition can be created by the noticing, the fluency enhancing, the hypothesis testing, the metalinguistic, the negotiating, and the generative processing function of an output task.
	Theoretical ground	Hulstijn & Laufer's Involvement Load Hypothesis (2001); Long's Interaction Hypothesis (1996); Nation's Vocabulary Generation (2013); Swain's Output Hypothesis (1985, 1995, 2005)
	Empirical ground	de la Fuente (2002), R. Ellis et al. (1994), R. Ellis & He (1999), R. Ellis & Heimbach (1997); Hulstijn & Laufer (2001); Kim (2008); Joe (1998); Loschky (1994); Lu (2013); Rassaei (2015); Yi & Sun (2013)

As all interventions in this research project have a common goal of fostering L2 learners' incidental vocabulary gain through facilitating their L2 listening comprehension, it might be pedagogically pertinent to know which intervention fares the best. To do so, a comparison of the effects on the size of vocabulary gain of these interventions must be carried out. This comparison is feasible thanks to the advent of systematic review or also known as meta-analysis (Ellis P., 2010). Such a comparison will be made in Study 6. In case there is indeed difference in the size of vocabulary gains across the interventions, different factorial analyses will also be conducted to search for explanatory factors.

It should also be noted that these experiments, due to their quantitative nature, cannot provide any direct evidence for the cognitive and metacognitive processes that learners may use to deal with the summary tasks in Study 3 and 5 and how these processes affect their text comprehension and incidental vocabulary acquisition. Thus, Study 7 will replicate Study 3 and 5, but with only eight learners. After the learners have completed their assigned experimental procedures, they will be asked to report on the aforementioned processes and the effects of these processes on their text comprehension and incidental vocabulary uptake.

CHAPTER 2: METHODOLOGY

In this chapter, I shall first state the research aims, research questions, and research designs for all studies. Next, the number of research participants, together with their prior L2 proficiency level and vocabulary knowledge, will be given in Section 2. In Section 3, I will focus on the way I have selected the input materials and the potential target words. The process of designing and validating dependent measures, including four vocabulary tests and two text comprehension tests, will be presented in Section 4 and 5. In the final two sections – 6 and 7, the procedures of collecting and analysing data will be described in detail.

Research Aims, Research Questions and Research Designs

As mentioned in the previous chapter, this research project includes five quantitative studies (Study 1 to Study 5), a comparison of the effect sizes of all treatment conditions on vocabulary gain (Study 6), and one qualitative study (Study 7). The research aim and research questions of each study are presented in Table 3:

Table 3: Research aims and research questions		
Study		
1	Research aim	To measure the effect of viewing a video once (henceforth referred to as <i>One-time Viewing</i> or <i>OV</i> for short) on the size of incidental L2 vocabulary uptake
	Research questions	(a) To what extent do L2 learners incidentally pick up the form and meaning of novel words from the OV condition? (b) Is this vocabulary gain associated with the learners' text comprehension? (c) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference?
2	Research aim	To measure the effect of viewing a video twice (henceforth referred to as <i>Repeated Viewing</i> or <i>RV</i> for short) on the size of incidental L2 vocabulary uptake
	Research questions	(d) To what extent do L2 learners incidentally pick up the form and meaning of novel words from the RV condition? (e) Is this vocabulary gain associated with the learners' text comprehension? (f) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference?

3	Research aim	To measure the effect of viewing a video, making an oral summary of the content, and then viewing this video again (henceforth referred to as <i>Repeated Viewing plus Summary</i> or <i>RVS</i> for short) on the size of incidental L2 vocabulary uptake
	Research questions	(g) To what extent do L2 learners incidentally pick up the form and meaning of novel words from the RVS condition? (h) Is this vocabulary gain associated with the learners' text comprehension? (i) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference?
4	Research aim	To measure the effect of viewing a set of three videos on similar topics (henceforth referred to as <i>Narrow Viewing</i> or <i>NV</i> for short) on the size of incidental L2 vocabulary uptake
	Research questions	(j) To what extent do L2 learners incidentally pick up the form and meaning of novel words from the NV condition? (k) Is this vocabulary gain associated with the learners' text comprehension? (l) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the videos, and the usefulness of verbal and pictorial clues in these videos for word-meaning inference?
5	Research aim	To measure the effect of viewing either the first or the second half of the three-video set in Study 4, sharing the content to a classmate who has not watched the same half yet in an interactional summary activity, and then moving on with the other half (henceforth referred to as <i>Narrow Viewing plus Summary</i> or <i>NVS</i> for short) on the size of incidental L2 vocabulary uptake
	Research questions	(m) To what extent do L2 learners incidentally pick up the form and meaning of novel words from the NVS condition? (n) Is this vocabulary gain associated with the learners' text comprehension? (o) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the videos, and the usefulness of verbal and pictorial clues in these videos for word-meaning inference?
6	Research aim	To compare the effect sizes of all learning conditions (OV, RV, RVS, NV, and NVS) on vocabulary gain
	Research questions	(p) Is there any difference in the effect sizes of all learning conditions on vocabulary gain? (q) If yes, which factor or factors in these conditions account(s) for this difference?

7	Research aim	To examine the cognitive and metacognitive processes that L2 learners use to deal with the summary activities in the RVS condition and the NVS condition and how these processes affect their text comprehension and incidental vocabulary acquisition
	Research questions	(r) To what extent do L2 learners incidentally pick up the form and meaning of novel words from the RVS condition and NVS condition? (s) What cognitive and metacognitive processes do they report using to perform the summary activities in these conditions and how do these processes affect their text comprehension and incidental vocabulary acquisition?

For the first five studies, I will use a classroom experiment with a mixture of both within- and between-participant research designs. To be more specific, two groups of L2 learners who have a similar level of L2 proficiency and vocabulary knowledge will be recruited for each study. One group will be a treatment group partaking in one of the learning conditions mentioned above (OV, RV, RVS, NV, or NVS). The other will serve as the control group. Vocabulary gain is gauged by a repeated pre-test – post-test – delayed post-test measure. Any score changes in the case of the control group will be regarded as the effect of their test-taking and chance learning of target words outside of the experiment. This will help me to determine to what extent the changes in test scores between the pre-test and the post-tests in the treatment group can be attributed to the corresponding treatment condition. The difference in the size of vocabulary gains between the groups (if any) will be used to compute the effect size of the corresponding treatment condition.

In Study 6, the effect sizes of all treatment conditions on vocabulary gain (as found from Study 1 to Study 5) will be integrated into a meta-analysis program to rank these conditions in order of their effect sizes. One advantage of meta-analysis is that it helps to calculate the true effect size of each learning condition by removing measurement errors (if any). In case there is indeed a difference in the effect sizes across these conditions, multiple regression analyses will be carried out in this study to determine which factor or factors in the learning conditions, the input materials or the learners explain(s) this difference.

Obviously, all studies above are quantitative in nature. One limitation of this is that they cannot provide direct evidence for *what actually happens in learners' mind, especially regarding the effect of the summary tasks in Study 3 and 5 on their incidental vocabulary acquisition*. A qualitative investigation is therefore added to this research project in order to overcome this limitation. In this investigation, eight L2 learners are recruited. Four of them are required to undertake the same procedure as those in the RVS condition in Study 3 while

the other four will follow the same procedure as those in the NVS condition in Study 5. When they finish their experimental procedures, I will have a semi-structured one-to-one interview with each of these learners about the cognitive and metacognitive processes they have used to deal with the output tasks and how these processes have affected their text comprehension and incidental vocabulary acquisition.

The ultimate goal of this research project is to explore different ways to facilitate learners' L2 listening comprehension and consequently enhance the likelihood of their incidental vocabulary uptake from the listening input. It is hoped that the findings from this research project will inform L2 course designers and instructors in their decision-making regarding how to develop meaning-focused listening activities that can foster not only L2 listening comprehension but also incidental vocabulary acquisition.

Research Participants

Participants in the quantitative studies (Study 1, 2, 3, 4, and 5) were 192 Vietnamese learners of English as a foreign language (six males and 186 females) recruited from a university in Vietnam (see Appendix C for their personal profiles). They were all 19 or 20 years of age and enrolled in a two-year intensive English language training program to develop their English language proficiency to CEFR² C1 level or IELTS³ overall band score of 6.5.

In their intact classes, these learners were assigned to the control condition ($N = 32$), the OV condition ($N = 32$), the RV condition ($N = 32$), the RVS condition ($N = 32$), the NV condition ($N = 32$), and the NVS condition ($N = 32$). Prior to this research project, their English language proficiency, as measured by a series of in-house English language proficiency tests using CEFR as the underlying test construct, was found to be equivalent to CEFR B2+ or IELTS overall band score of 5.5 (i.e., upper intermediate). Due to the confidentiality of these tests, further information on these tests could not be disclosed in this thesis. Scores on Nation and Beglar's (2007) Vocabulary Size Test (VST) indicated that these learners had receptive knowledge of, at least, the 4,000 most frequent word families in English. As all the data were normally distributed, two one-way ANOVA tests were carried out to compare the learners' L2 proficiency scores and VST scores across the six groups. No difference was found either in the case of their L2 proficiency scores or in the case of their

² CEFR: The Common European Framework of Reference for Languages

³ IELTS: International English Language Testing System

VST scores, with $F(5, 186) = 0.61$ ($p = .69$) and $F(5, 186) = 1.24$ ($p = .29$), respectively. This means that the learners in all groups had a roughly equal level of L2 proficiency and vocabulary knowledge before they were involved in this research project.

Participants in the qualitative investigation (Study 7) were eight Vietnamese learners of English as a second language recruited from a university in New Zealand (also refer to Appendix C for their personal profiles). They were all female post-graduate students. Their ages ranged from 27 to 36 with a mean of 31.63 (SD 2.93). Prior to this study, their L2 proficiency, as measured by one of the official IELTS tests, ranged from 6.5 to 7.5 overall band scores with a mean score of 7.13 (SD 0.44) and their VST scores were between 6,000 and 7,500 most frequent word families in English with a mean of 6,987.50 (SD 508.32). Obviously, these learners had a noticeably higher level of both L2 proficiency and vocabulary knowledge than those in the five quantitative studies above. It should be, however, noted that the primary aim of this qualitative study is not to compare the size of vocabulary gain obtained by these learners with those in Study 3 and 5, but to examine the cognitive and metacognitive processes that the learners used to deal with the output tasks and how these processes affected their text comprehension and incidental vocabulary uptake. Therefore, the difference in their L2 proficiency and vocabulary knowledge between the studies does not compromise the validity of this investigation or the research project as a whole.

All participants were informed about (a) the general purpose and experimental procedures of this research project (except for the presence of the vocabulary tests), (b) the benefits and responsibilities they had when involved in this research project, and (c) all relevant ethical issues before they decided whether or not to participate in this research project by signing a consent form (see Appendices A and B for the information sheet and the consent form).

Input Materials and Potential Target Words

Input Materials

TED Talks were chosen as the input materials in this research project. These talks are real presentations at the Technology-Entertainment-Design (TED) Conference, which is annually held for one week on the West Coast of North America. During this week, over 70 experts from a wide range of academic and professional disciplines come and share their ground-breaking ideas to a group of lay and semi-lay audiences in a manner that the

conference attendees often describe as “*the ultimate brain spa*” (this description was taken from the conference official website <https://www.ted.com> on July 07 2016). English is used as the sole medium of communication in this conference. Obviously, these talks are originally created for real life rather than L2 instruction purposes. Therefore, they can be used as authentic listening materials in L2 courses in this regard (Coxhead & Walls, 2012; Nurmukhamedov & Sadler, 2011). As this research project aims to gauge incidental L2 vocabulary uptake from authentic listening materials, TED Talks meet this requirement.

Takaesu (2013) integrated TED Talks into a L2 course in a college in Japan. In the end-course survey, over two thirds of the surveyed learners ($N = 303$) reported that TED Talks indeed helped them to substantially improve their L2 listening ability, enhance their L2 listening confidence, and familiarize them with a wide variety of English accents. It should be also noted that such benefits of TED Talks for L2 listening were not only found in the case of proficient but less-proficient learners as well. This study therefore confirmed the claims made by Coxhead and Walls (2012) and Floyd and Jeschull (2012) that TED Talks have merits for L2 listening development. In this research project, I am looking for videos which can foster L2 listening comprehension, and so TED Talks appear to fit the bill.

Coxhead and Walls (2012) created a corpus of 60 six-minute TED Talks videos with a total number of 43,656 running words. This corpus was then run through Nation and Heatley’s Range Program (2002) to get the frequency-based rankings of these word families. The results showed that with a vocabulary size of 4,000 most frequent word families plus proper nouns in English as in the case of the participants in this research project, learners can reach the 95% lexical coverage or proportion of known words in these talks. This lexical coverage is generally believed to enable learners to generate a sufficient understanding of the talks (Bonk, 2000; Schmitt, 2008; van Zeeland & Schmitt, 2012) and thereby to successfully infer the meaning of novel words from the surrounding context (Nation, 2013; Nation & Webb, 2011). Out of the 43,656 running words, 4.53% are beyond the 4,000 most frequent word families. This means that about 33 words per six-minute TED Talks video might be unknown to the participants in this research project and therefore available to measure incidental uptake.

Most of TED Talks aim to deliver specialized knowledge. However, Scotto di Carlo (2014) found from his analysis of 80 TED Talks sampled from various disciplines that whenever it comes to technical concepts or complex ideas, the presenters in these talks often

use everyday metaphors, comparisons, and/or exemplifications to explain these ideas in a highly intelligible way. Caliendo and Compagnone (2014) also found that to compensate for the lack of interaction and therefore opportunities for negotiation of meaning between the speaker and the audience in such a talk, the speaker tries to organize the intended message in a highly structured fashion with an effective use of metatalk or discourse markers. Previous research has furnished evidence for the usefulness of these discourse features for both L2 listening comprehension and incidental vocabulary acquisition (e.g., Jung, 2003; Morrel, 2007; Sawada, 2009; Seo, 2002; Vidal, 2003, 2011).

Regarding practicality issues, there were, at the time of writing this thesis, more than 2200 TED Talks videos available at the conference official website (this figure was also taken on July 07, 2016). These videos were categorized under six major themes: Technology, Entertainment, Design, Business, Science, and Global Issues. Each theme included many different sub-topics. The total number of existing topics amounted to 400 (this figure was also taken on July 07, 2016). L2 learners thus stood a good chance of finding a topic of their own interest. These talks typically lasted from four to twenty minutes. Therefore, they were manageable for classroom use. Finally, they were totally free to access and download either in the audio format (MP3) or the video (MP4) format. This helped to avoid copy-right issues.

The review above indicates that TED Talks can be used as authentic audio-visual input materials in L2 courses. These talks also have great potential for fostering both L2 listening comprehension and incidental vocabulary uptake. Therefore, they were felt suitable as the input materials in the current research project.

Sixteen learners who had a similar level of L2 proficiency and vocabulary knowledge as those in the quantitative studies in this project were invited to rate samples of TED Talks videos on various topics with regard to their comprehensibility and general appeal. Based on the learners' feedback, three videos were chosen. These videos are all about ocean wonders and last five, eight and twelve minutes, respectively (please refer to Appendix D for further information on these videos and their transcripts). The three videos will be used for Study 4, 5 and 7. However, only the longest video (i.e., the twelve-minute one) will be used for Study 1, 2, 3 and 7. Another three videos that have the same total length of 25 minutes as the three videos above, but are about different topics (i.e., "*eight secrets of success*", "*leather and meat without killing animals*", and "*this app knows how you feel*") and include no target words (i.e., words selected to gauge vocabulary acquisition from watching the videos about ocean

wonders) were selected for use in the control condition. Henceforth, these two video sets will be referred to as the experimental and the non-experimental videos, respectively.

In the previous chapter, I have mentioned that *Narrow Viewing* is believed to help learners to increase their lexical coverage when they move from one video to another. This benefit is now weighed for the case of the three experimental videos by using Nation and Heatley's Range Program (2002). The results show that, if they were to view just one of these videos, the participants in this research project, given their estimated receptive knowledge of at least 4,000 most frequent word families in English, will have a lexical coverage of 95.90%, 97.39%, and 95.66% for the five-, eight- and twelve-minute video, respectively. In other words, there are 4.10%, 2.61% and 4.34% of the word families in these videos that might be unfamiliar to these participants. However, there are four overlapping unfamiliar word families between the five- and the eight-minute video. Thus, if the participants view the five-minute video before the eight-minute one, *and* manage to develop knowledge of these unfamiliar words, then they can increase their lexical coverage for the latter video from 97.39% to 98.50%. There are also eleven overlapping unfamiliar word families between the combination of the five- and the eight-minute video on the one hand and the twelve-minute video on the other hand. Therefore, if the participants view the three experimental videos in a sequence from the shortest to the longest one, they might in theory increase their lexical coverage of the longest video from 95.66% to 97.03% - provided incremental acquisition occurs through encountering the words. In accordance with the proclaimed benefits of narrow viewing, viewing the three experimental videos in succession can thus be expected to reduce the lexical burden for the participants as they tackle the third video.

I opted for the uncaptioned version of these videos because this research project examines vocabulary gain from listening, not reading. When captioned videos are used, it is impossible to tell whether vocabulary gain occurs as a result of listening, reading the captions, or both. Previous research suggests that L2 learners allocate a lot of their attention to captions (Montero Perez et al., 2015; Winke et al., 2013) and rely on their better developed reading skills to bypass the need to process the listening input (Garza, 1991). In any case, the question of whether captioned videos actually foster incidental L2 vocabulary acquisition better than non-captioned videos is still not settled. Although several studies have reported that viewing captioned videos leads to a significantly larger incidental L2 vocabulary gain than viewing the same videos without captions (e.g., Markham, Peter & McCarthy, 2001; Montero Perez et al., 2013; Neuman & Koskinen, 1992; Sydorenko, 2010; Winke, Gass, &

Sydorenko, 2010), other studies have contradicted this (e.g., Etemadi, 2012; Rodgers, 2013; Yuksel & Tanriverdi, 2009). Besides, as already mentioned, many studies that have claimed superiority of captioned over non-captioned videos for vocabulary learning used written L2 word prompts in post-tests regardless of whether participants watched non-captioned materials, which arguably introduces a confounding variable due to input-modality – test-modality (in)congruency (Sydorenko, 2010).

Potential Target Words

The Range Program was also used to identify potential target words for this research project. From the lexical profile of the longest experimental video, 20 words that are beyond the most frequent 4,000 word families in English were chosen as the potential target words for Study 1, 2, 3, and 7. Together with these 20 lexical items, another 20 lexical items that are also beyond the 4,000 most frequent word families in English were selected from the other experimental videos as the potential target words for Study 4, 5 and 7. Choosing these words reduces the likelihood that the participants might already know these lexical items prior to the experiment. The potential target words for Study 1, 2, 3, and 7 and for Study 4, 5, and 7, together with their parts of speech and the frequency of their occurrences in the input materials, are presented respectively in Table 4 and 5 below:

Table 4: Potential target words for Study 1, 2, 3, and 7

Words	Part of speech	In-text occurrence	Words	Part of speech	In-text occurrence
Bait	Noun	2	Mesh	Noun	2
Barbel	Noun	2	Morsel	Noun	1
Barrage	Noun	2	Plankton	Noun	2
Bioluminescence	Noun	6	Plume	Noun	2
Breath-taking	Adjective	2	Plunge	Noun	2
Contemplative	Adjective	1	Propagate	Verb	1
Entranced	Adjective	1	Saber	Noun	1
Foolhardy	Adjective	1	Trawling	Noun	1
Junky	Noun	1	Twinkling	Adjective	1
Lure	Noun	1	Unobtrusive	Adjective	1

Table 5: Potential target words for Study 4, 5 and 7

Words	Part of speech	In-text occurrence	Words	Part of speech	In-text occurrence
Bait	Noun	2	Lure	Noun	1
Barbel	Noun	2	Mesh	Noun	2
Barrage	Noun	2	Morsel	Noun	1
Bioluminescence	Noun	6	Outrageous	Adjective	1
Bluffing	Verb	1	Plankton	Noun	2
Breath-taking	Adjective	2	Plume	Noun	2
Camouflage	Noun	1	Plunge	Noun	2
Colonial	Adjective	1	Propagate	Verb	1
contemplative	Adjective	1	Puddle	Noun	1
Crevice	Noun	1	Pulsating	Adjective	1
Droopy	Adjective	1	Reef	Noun	1
Entranced	Adjective	1	Saber	Noun	1
Firefly	Noun	2	Snorkel	Noun	1
Flailing	Verb	1	Sonar	Noun	2
Foolhardy	Adjective	1	Spook	Verb	1
Footage	Noun	1	Startling	Adjective	1
Hull	Noun	1	Trawling	Noun	1
Hypnotize	Verb	1	Twinkling	Adjective	1
Joystick	Noun	1	Unobtrusive	Adjective	1
Junky	Noun	1	Vampire	Noun	2

Regarding the usefulness of verbal clues in the input materials for word-meaning inference, two experienced teachers of English as a foreign/second language (EFL/ESL) (one a native and the other a non-native speaker of English) were first invited to read the transcripts of the videos, stop whenever they met with a target word, and then rate how useful the surrounding verbal clues were for them to infer the meaning of this word on a five-point scale from “not very useful” (1) to “very useful” (5). For example, the target words “**junky**” and “**bioluminescence**” in the following contexts “*Since my, my first dive in a deep diving submersible, when I went down and turned out the lights and saw the fireworks displays, I have been a bioluminescence **junky**. But I would come back from those dives and try to share the experience with words, and they were totally inadequate to the task. I needed some way to*

share the experience directly.” and “*So it is a little appreciated fact that most of the animals in our ocean make light. I have spent most of my career studying this phenomenon called bioluminescence. I study it because I think understanding it is critical to understanding life in the ocean where most bioluminescence occurs.*” were rated as “*not very useful*” (1) and “*very useful*” (5), respectively. For any word with more than one time of in-text occurrence, the teachers were informed beforehand to rate the usefulness of these clues any time they met this word again. The mean of their ratings was used for the word. The independent ratings of the teachers were also correlated with my own ratings. The correlation coefficients in the case of the 20 potential target words are .80, .83, and .88 ($p < .0001$). These coefficients in the case of the 40 potential target words are .85, .87, and .82 ($p < .0001$). The mean ratings of the three independent raters, including mine, will be used for the final analyses.

For the usefulness of pictorial clues in the videos for word-meaning interpretation, these teachers were first required to watch an English captioned version of the videos, stop whenever they encountered or reencountered one of the target items, and then rate how useful the surrounding pictorial clues were for them to decipher the meaning of this word on a five-point scale from “*not very useful*” (1) to “*very useful*” (5). Again, if this word occurred more than once in the videos, the mean of all relevant ratings was used for this word. The independent ratings by these teachers were subsequently correlated with my ratings. The correlation coefficients are .83, .88, and .81 ($p < .0001$) in the case of the 20 potential target words and .81, .83, and .87 ($p < .0001$) in the case of the 40 potential target words. The mean ratings of the three independent raters, including mine, will be used for the final analyses.

It should be noted here that all the words above are the *potential* target items in this research project. The actual target words will be selected based on the participants’ actual pre-test scores. In addition, these potential lexical items are chosen mainly because they are unlikely to be known to the research participants prior to the experiments rather than because they are deemed to have particular value or relevance for these learners. This is a limitation of this research project.

Dependent Measures in Study 1, 2, 3 and 7

These studies have the same dependent measures: a text comprehension test and two vocabulary knowledge tests. In what follows, I will present the way these tests have been designed and validated.

Text Comprehension Test

Selecting and operationalizing the test construct

This test aims to measure the extent to which the participants understand and remember the content of the longest video and the test scores will be used once and for the purposes of those studies only. Following Bachman and Palmer (1996, p.118), its corresponding test construct should be theory-based rather than syllabus-based. From the current theories of L2 listening comprehension in general (Buck, 2001, pp.02-03; Flowerdew & Miller, 2005, pp.24-27; Nation & Newton, 2009, pp.40-42) and L2 academic listening comprehension in particular (Flowerdew, 1994, pp.09-13), scholars in this area have developed a two-pronged model of top-down and bottom-up processing. In this model, it is theorized that L2 learners use both top-down and bottom-up processing, but in different proportions for different listening purposes. In top-down processes, they use their previous linguistic and background knowledge to form, test, and then either accept, modify or reject their hypotheses about the listening content. In bottom-up processes, they construct a mental representation of the given information by accretion or, in other words, combining increasingly larger units of meaning from the phoneme level all the way up to the discourse level (Vandergrift, 2004).

To the best of my knowledge, Wagner (2002) is the first and the only one to date that has assessed the validity of this model. In this study, 85 learners of English as a second language watched three academic videotexts and then completed 20 content questions that were deemed to measure either their top-down or bottom-up processing ability. Using a series of exploratory factor analyses, Wagner indeed succeeded in providing statistical evidence for the validity of this model. Therefore, this model is used to define the construct of the current test.

To operationalize this test construct, it is worth revisiting the tasks that previous research has recommended as relevant and representative of L2 academic listening (Bachman & Palmer, 1996, pp.102-107; Flowerdew, 1994, pp. 10-20). Recently, Kang et al. (2016) have made a comprehensive review of 20 validated L2 academic listening tests, and then built up an aggregate taxonomy of the relevant and representative tasks to assess this listening ability. From this taxonomy, the two most important tasks to operationalize the top-down processing are to require L2 learners to (a) identify the topic and main ideas of a given listening passage and (b) make inferences about its implicit meaning, logical links between ideas, and speaker

attitude towards that topic. Meanwhile, the bottom-up processing is often operationalized by asking L2 learners to (c) recognize its supporting ideas or factual information. Therefore, it seems justified to use the three tasks above to operationalize the underlying construct of the current test.

Designing the test

In this stage, I need to decide the specific question types for the three tasks above. In an attempt to validate the cognitive processes prompted by the current question types (i.e., gap-filling and multiple choices) in the standardized IELTS academic listening comprehension section, Field (2012) compared the ways in which L2 test-takers cognitively processed academic input texts in a real-time condition and in an IELTS testing condition. The results showed that, in the testing condition, the gap-filling questions tended to direct the test-takers' attention towards the word level at the expense of the sentence and discourse level. Meanwhile, the multiple-choice questions required too much reading, which consequently violated the underlying construct of listening comprehension. A better question type that Field recommended for such an academic test was Yes/No/Not Given Statement (p.437), which is therefore used for the present measure.

In fact, this question type has many other benefits, which are relevant to the purpose of the current test. First, it allows the measurement of a large amount and various types of information in a short period of time (Rodgers, 2013). Second, it entails a reliable measurement in the sense that it involves little or no subjective scoring. Finally, L2 learners have long been familiarized with this question type (Frisbie & Becker, 1991; Haladyna, 1999; Kreiter & Frisbie, 1989). As a result, its face validity and response validity is likely to be optimized.

However, in order not to put too much cognitive demand on the participants in the current test, the options "No" and "Not Given" are merged as "No" in the design process. On the downside, this merge increases the chance the participants make a correct choice to a test item merely by random guessing from 33.3% to 50%. Whether such guesses occur and the extent to which they (if any) affect the quality of the test as a whole will be examined in the test validation process that I am going to report further below.

To prepare the test content, three experienced EFL/ESL teachers (one a native and the two others non-native speakers of English) were invited to watch the video and then read its

transcript to identify its topic, main ideas, subtle, but yet important supporting ideas, factual information, implicit meaning, logical link between ideas, and speaker attitude towards the topic. Fifteen idea units that all teachers listed were used to design 15 Yes/No questions. These questions were constructed by either paraphrasing the information presented in the video for “Yes” items or adding contradictory/irrelevant information for “No” items. All the words used in these items are within the most frequent 3000 word families in English in order to ensure that the participants (given their estimated vocabulary size of at least 4000 most frequent word families in English) would have little trouble in understanding the test prompts. After these items were drafted and piloted three times with three different groups of EFL/ESL learners, the native English-speaking teacher above was invited to do any necessary text editing. This edited/revised version was used in Study 1, 2, 3 and 7 (please see Appendix H). In the scoring procedure, 1 and 0 point were awarded to any correct and incorrect answers, respectively. In the following validation, the test scores obtained by all the participants in the treatment group in Study 1 were used ($N = 32$).

Validating the test quality

This validation uses Messickian construct validity and Rasch Model Analysis as a theoretical framework and analytical tool, respectively. According to Messick (1995), a language test is considered valid or useful for its purpose if it satisfies all the criteria in the content, substantive, structural, generalizable, external, and consequential aspect. Rasch Model is specially designed to bring statistical evidence for such a validation (Baghaei & Amrahi, 2011). It should be, however, noted that the current test will be used once and for the purposes of those studies only. Therefore, this test is unlikely to have a sizable impact on the participants’ formal education. That is why, in this validation, only evidence related to the first five aspects, but not the consequential aspect, will be gathered and scrutinized.

Content aspect

The content aspect of Messickian construct validity is concerned with the relevance and representativeness of test content to the test construct and the technical quality or also known as the quality of item writing, key system and scoring procedure. In the following section, the insights of these criteria will be first introduced, and then their corresponding evidence from Rasch Model outcome will be used to evaluate the degree to which the test satisfies these criteria.

The extent to which test content is relevant to its underlying construct is often assessed through the way a test construct has been defined and operationalized and the way the test items have been designed to fit this construct. To facilitate this evaluation, I need to provide in my test specification a clear description of the test purpose, test construct, operationalization of this test construct, and procedure of designing the test items (Bachman & Palmer, 1996, pp.95-133). All this information has been presented above.

To provide evidence for the representativeness of the test content, I first need to ensure that the difficulty levels of the current test items are sufficient to measure what the test intends to measure or, in other words, the text comprehension levels of the participants. Rasch Model provides this evidence through its Person-Item Map. To construct this map, this model estimates the text comprehension level of each participant and the difficulty level of each test item from the provided raw test scores. Subsequently, it arranges these estimated values along the two sides of the same continuous scale, using logits as a common measurement unit. Contrasting the range of the text comprehension levels against that of the item difficulty levels, I am able to decide whether the difficulty levels of the current test items adequately measure the text comprehension levels of the participants. For example, if the estimated values of all text comprehension levels in the current case are between -1.00 and +3.00 logits and those of all item difficulty levels range between -2.13 and +5.00 logits or, in other words, the range of the text comprehension levels falls neatly within the range of the item difficulty levels, the difficulty levels of the current test items effectively measure the text comprehension levels of the participants. However, if the estimated values of all text comprehension levels range from -1.00 to +3.00 logits but those of the item difficulty levels go from -1.00 to +2.00 logits, more test items are required to fill in the gap from +2.00 to +3.00 logits.

The other crucial evidence for content representativeness is the degree to which the difficulty levels of the current test items are able to discriminate the text comprehension levels of the participants (or also known as “item discrimination” in the conventional item analysis theory). Rasch model provides this evidence through an index called “item strata” or “the number of statistically distinct regions of item difficulty that the persons [the participants] have distinguished” (Smith, 2001, p.293). Smith (2001) proposes item strata of at least 2.00 as an indicator for an acceptable level of item discrimination.

Regarding the technical quality, I need to provide two different types of statistical evidence, which can be taken directly from the Rasch Model outcome: (a) item infit mean squares and the corresponding item fit statistics (henceforth referred to as ZSTDs for short) and (b) item-measure correlation coefficients. While the former indicates whether a particular test item fits with the underlying test construct, the latter shows whether it behaves consistently with all the other items in the same test. Baghaei and Amrahi (2011) suggest a combination of (a) item infit mean squares from 0.70 to 1.30 logits, the corresponding ZSTDs from -2.00 to 2.00 logits, and (b) item-measure correlation coefficients above 0.00 as indicators for an acceptable technical quality.

Let me now come back to the case of my current test, the degree to which the test content was relevant to the selected construct could be evaluated through the process of selecting and operationalizing the test construct and the process of designing the test items to fit this construct as described above. In what follows, I will only report the evidence for the representativeness of the test content and the technical quality. Figure 1 presents the estimated values of the text comprehension levels of all participants (in Study 1) ($N = 32$) and the difficulty levels of all test items ($N = 15$) on the upper and lower chart, respectively:

Figure 1: Person-item map

As can be seen from the Rasch Model Item-Person Map in Figure 1, the estimated values of the text comprehension levels of 32 participants ranged from -1.50 to +2.27 logits whereas those of the difficulty levels of 15 test items were from -2.31 to +5.06 logits. There was a serious gap in the hierarchy of the item difficulty levels (at the region of +2.0 logits), suggesting that more test items were required to fill in this gap. However, it should be noted that this gap was associated with the text comprehension level of only one participant. The

data given by this participant will be removed from the final analyses. The item strata were 2.26, which is larger than Smith's (2001) required item strata of 2 as an indicator for an acceptable item discrimination level. The internal reliability of the whole test was .79. Regarding the technical quality, the item infit mean squares and the corresponding ZSTDs were between .74 and 1.12 and between -1.90 and 1.20, respectively. The item-measure correlation coefficients ranged from .16 to .83. All these indices fell neatly within the acceptable ranges for an acceptable technical quality (Baghaei & Amrahi, 2011).

It is therefore evidenced that the current test included only relevant and representative content for the selected underlying construct. The test prompt was clearly written and included an acceptable amount of reading. No problem was found either for the key system or the scoring procedure. In other words, this test satisfied all the criteria for the content aspect of Messickian construct validity.

Substantive aspect

In order to make a valid inference about L2 academic listening abilities of the participants from their test scores, I need to ensure that, in their responses to all test items, the participants have indeed exercised the cognitive processes which the test items intend to measure. However, this is not always the case in reality. For instance, a participant may make a careless mistake or a random, but correct guess in their response to a particular test item. Therefore, the test score on this item is not a valid indicator of his/her true academic listening ability. This is basically what Messick (1995) refers to as the substantive aspect of his construct validity.

As already mentioned above, Rasch Model estimates both text comprehension and item difficulty levels in logit values. In this way, I am able to compare the text comprehension levels across all participants and the difficulty levels across all test items. For example, a participant with a text comprehension level of +2.0 logits proves to be more successful on this measure than another participant with a text comprehension level of -1.0 logit. Likewise, a test item with a difficulty level of +4.0 logits proves to be much more challenging than another item with a difficulty level of -2.0 logits. Based on these estimated values, Rasch Model provides statistical evidence for the substantive aspect through detecting the participants with unexpected item responses. There are two inter-related underlying assumptions that guide Rasch Model in this detection. On the one hand, a participant with a high total test score is expected to succeed in an easy test item. On the other hand, a

participant with a low total test score is likely to fail on a difficult test item. Therefore, if a participant with a high total test score gives a wrong answer to a very easy question or a participant with a low total test score gets a correct response to a very difficult item, either the item or the participant is problematic or “*misfit*”. Rasch Model reveals these misfit participants through their outfit mean squares and the corresponding ZSTDs in the model outcome. Specifically, these misfit participants are those with the outfit mean squares and the corresponding ZSTDs larger than 2.00 (Baghaei & Amrahi, 2011).

It is advised that the data provided by these participants should be first removed from the data matrix and Rasch Model analyses should be re-run for the content aspect (Rodgers, 2013). If there is a significant improvement in the quality of the content aspect, especially in its content representativeness and technical quality, it is these misfit participants that have distorted the quality of the test content or the test items. Otherwise, the items that these participants have given unexpected responses to are indeed misfit and, thus, should be considered for revision or exclusion from the test.

In the current test, there were two participants whose actual text comprehension levels were not matched with the model expectation. It is evidenced by their outfit mean squares (+4.24 and +4.13 logits) and the corresponding ZSTDs (both were +2.2 logits), which are all above +2.00 logits. Their unexpected responses to some test items might have distorted the quality of these test items. To further examine this issue, the data provided by these participants were first removed from the data matrix. Rasch Model analyses were then implemented again. However, no significant improvement was found in the quality of the content aspect. This suggests that although these participants produced unexpected responses to some test items, this did not compromise the quality of the test as a whole. Given the small sample size of this research project, the data provided by these test-takers were all retained.

Alternatively, I can bring evidence for the substantive aspect of Messickian construct validity by testing the hypotheses that I have set for the difficulty levels of all test items and the text comprehension levels of all participants in the process of test design (McLean, Kramer & Beglar, 2015). In the current test, three types of information have been used to construct the test items, as follows: (a) the purpose and main ideas of the video, (b) its subtle, but yet important supporting ideas and key factual information, and (c) its implicit meaning, logical link between ideas and speaker attitude towards the talk topic. In his validation of a L2 academic listening test, Wagner (2002) found that his L2 test-takers scored significantly

better on the items related to the purpose and main ideas of the listening materials than the items related to the supporting ideas and factual information. Their scores on these two types of test items, in turn, significantly outweighed those on the test items related to implicit meaning, logical link between ideas or speaker attitudes towards the given topics. Therefore, this finding has been used to form the first hypothesis for the item difficulty levels in my current test.

As not all data are normally distributed, I resorted to a Kruskal Wallis test as a non-parametric counterpart of an ANOVA test for an independent-sample comparison. A significant difference was found in the test scores across the three types of test items: $X^2(2, N = 96) = 79.57$ ($p < .0001$). Conover tests were carried out for the pairwise comparisons with p values adjusted according to the family-wide error rate procedure of Holm. The results showed that the scores on the test items related to the talk purpose and main ideas were indeed significantly higher than those on the items related to the supporting ideas and factual information ($p < .0001$), which, in turn, were significantly higher than the scores on the items related to implicit meaning, logical link between ideas, and speaker attitude towards the given topic ($p < .0001$).

In addition, it was also hypothesized in the current test that the “No” items were more challenging than the “Yes” items as the participants normally focused on what was said rather than what was not said during their viewing session. Therefore, the test scores on the former were likely to be lower than those on the latter. As the dataset for this comparison is normally distributed, a t -test was used. The result indeed supports this hypothesis: $t = -4.11$ ($p < .000$). Following the previous research findings (e.g., Bonk, 2000; Stæhr, 2009, 2008; van Zeeland & Schmitt, 2013), it was also hypothesized that the larger L2 vocabulary knowledge the participants had prior to the experiment, the higher the text comprehension levels they tended to obtain. A Pearson correlation coefficient was computed for this association. Indeed, a positive coefficient of .45 ($p < .05$) was found.

Together, most of the participants were mentally engaged with the current test as expected. Though the test performance of two participants was not matched with the model expectation for their L2 academic listening abilities, this did not compromise the quality of the test as a whole. In addition, my initial hypotheses about the difficulty levels of the test items and the text comprehension levels of the participants were all statistically supported.

Therefore, this test also satisfied the criteria for the substantive aspect of Messickian construct validity.

Structural aspect

This aspect is concerned with whether a language test has only one underlying construct or, in other words, whether it is unidimensional. Linacre (2007) suggests that a language test is primarily unidimensional if the total variance in the test scores explained by the test itself is larger than 60% and the score variance in each residual component unexplained by this test is smaller than 3%. Rasch Model provides these indices through its Principal Components Analysis of test item residuals.

The current test was found to account for only 42.6% of the variance in the test scores. There were five residual components in its data matrix. In these components, the proportions of the score variance that could not be explained by the test ranged from 6.5% to 12.4%. All of these figures do not meet the thresholds that Linacre (2007) suggested as indicators for item unidimensionality. Thus, there might be a second meaningful construct in the test.

To further investigate this issue, Stevens' (2002) guidelines were used. In all the residual components, however, there were fewer than three, four and ten items with their absolute loadings over .80, .60 and .40, respectively. Therefore, the hypothesis that the current test had a second meaningful construct was statistically rejected. In other words, this test also met the requirement for the structural aspect.

Generalizable and external aspect

The generalizable aspect of Messickian construct validity deals with the extent to which the score meanings and interpretations are generalizable to other tasks or contents not included in a language test (due to the practicality issues), but belong to the same underlying construct (Baghaei & Amrahi, 2011). To check the generalizability of the current test, its items were first divided into two halves according to their odd and even numbering. Rasch Model analyses were subsequently run for each half of the test to generate the estimated values for the text comprehension levels of all participants. The disattenuated correlation or the correlation of their text comprehension levels between the two halves of the test *after removing the measurement error* was calculated by the formula below:

$$\text{Disattenuated correlation} = \frac{\text{Reliability index of the whole test}}{\sqrt{\text{Reliability index of one half} \times \text{Reliability index of the other half}}}$$

The disattenuated correlation in this case was 1.00, which was larger than the required disattenuated correlation of at least .90 that McLean, Kramer and Beglar (2015) proposed as an indicator for item invariance.

The external aspect is concerned with the degree to which the scores from the current test are correlated with those from another validated test or assessment practice (Messick, 1996). To validate the external aspect of the current test, the model estimates of the text comprehension levels of the participants were correlated with their scores on Nation and Beglar's (2007) VST, which was successfully validated by Beglar (2009). Although this correlation coefficient was relatively low ($r = .39, p < .05$), it was both significant and positive. Therefore, the test also had the external reliability.

Altogether, the current test successfully met the requirements for the content, substantive, structural, generalizable and external aspect of Messickian construct validity. Therefore, the test scores can now safely be used for further analyses.

Vocabulary Knowledge Tests

Selecting and operationalizing the test construct

As informed by the Literature Review Chapter, I have adopted Nation's model of L2 vocabulary knowledge (2013, p.50) to define the construct of these vocabulary knowledge tests. Receptive and productive knowledge of the most important aspect of vocabulary knowledge (Nation, 2001, 2013; Read, 2000) – form-meaning connection – was used to operationalize this construct.

Designing the tests

To design these tests, I followed Read's framework for L2 vocabulary assessment (2000, p.9). In this framework, Read specifies different options in developing a L2 vocabulary test as can be seen in Table 6 below:

Table 6: Read's dimensions of L2 vocabulary assessment (2000, p.09)

Discrete:				OR	Embedded:
A	measure	of	vocabulary		A measure of vocabulary which forms part of the

knowledge or use as an independent construct

assessment of some other larger construct

Selective:

A measure in which specific vocabulary items are the focus of the assessment

OR **Comprehensive:**

A measure which takes account of the whole vocabulary content of the material (reading/listening tasks) or the test-taker's response (writing/speaking tasks)

Context-independent:

A vocabulary measure in which the test-taker can produce the expected response without referring to any context

OR **Context-dependent:**

A vocabulary measure which assesses the test-taker's ability to take account of contextual information in order to produce the expected response

As I have selected Nation's trait definition of vocabulary knowledge as a theoretical basis to inform the design of my vocabulary knowledge tests, this knowledge should be measured independent of context (Chapelle, 1998). The main purpose of these tests is to examine whether the participants incidentally pick up the form and meaning of the 20 target words after their viewing sessions. Therefore, the tests should be "discrete" and "selective". The receptive and productive dimensions of this knowledge are gauged via a word-meaning recall test and a word-form recall test, respectively.

To prepare the word-meaning recall test, a male native speaker of English was invited to read aloud and audio-record the 20 potential target words with an interval of 20 seconds between the words. In the actual test, the participants are required to provide the meaning of the target words in Vietnamese (L1), prompted by the audio-recorded English word form. The test instruction and an example are presented in Vietnamese (L1) on the response sheet (see Appendix F and G for the test paper and a sampled test response, respectively). To prepare the word-form recall test, the male native speaker of English read aloud and audio-recorded the first or first two syllables of these words also with an interval of 20 seconds between them. In the actual test, the participants are asked to say each word aloud and audio-record it by an audio-recorder or smartphone, prompted by (a) its meaning written in Vietnamese on the test sheet and (b) its first or first two syllables presented in an audio-recording. In the case of monosyllabic target words, the ending sounds (or codas) (e.g., /t/ as in "bait" or /m/ as in "plume") are missing from the prompts. An experienced Vietnamese teacher of English checked the accuracy of L1 translations of the target words. During this

word-form recall test, the participants are required to wear an over-ear headphone with a microphone attached. This device helps them to listen to the test prompts and audio-record their responses without being distracted. Only four participants take the test at a time and they are seated at different corners of a large classroom. It takes the participants 15 minutes to complete each of the tests.

The word-meaning recall format has been successfully used in the studies by van Zeeland and Schmitt (2013) and Brown et al. (2008), for instance. However, the tests in these studies share a limitation in that all test prompts were presented in written form even though their research participants were not exposed to this form in the input materials. Sydorenko (2010) and Winke et al. (2010), for example, have found a significant difference in the vocabulary test scores when the presentation mode of the test prompts was either congruent or not congruent with that of the input text. Therefore, my current test uses the spoken rather than written word form as the prompts for the word meaning recall. Previous research (e.g., Sydorenko, 2010; van Zeeland & Schmitt, 2013) has already examined the extent to which L2 learners incidentally acquire the form of novel words through listening/viewing. However, their tests only focus on the participants' ability to recognize rather than recall the word form. In this study, I examine whether the participants are able to recall the form of novel words after their viewing sessions. The use of L2 spoken first or first two syllables is to confine responses to the target words rather than possible synonyms.

If the learners in this study were given both of these vocabulary tests, their exposure to the test prompts in one test might prime their responses in the other. Specifically, if they took the word-form recall test first, they might familiarize themselves with the target-word meaning before taking the word-meaning recall test. Conversely, if they were provided with the word-meaning recall test first, their exposure to the L2 word prompts would help them in the word-form recall test. To avoid such inter-test effects (which would make it impossible to attribute different success rates to the treatment conditions that participants have been assigned to), the word-meaning recall test in this study is assigned to half of the participants in each treatment condition while the word-form recall test is given to the other half. As a result, for the vocabulary measures, there are two different datasets (one for the word-meaning recall test and the other for the word-form recall test).

These tests were given to the participants three times: two weeks before the experiment, at the end of the experiment, and two weeks after the experiment (henceforth

referred to as the pre-test, the immediate post-test and the delayed post-test, respectively). Each time the tests were administered, the order of the test items was randomized to minimize test-item-sequence effects. In the scoring procedure of these vocabulary tests, 1 and 0 point were awarded to any correct and incorrect responses, respectively. For example, such responses as “*mỗi câu ca*” (“*bait*”) and “*đánh đập*” (“*beat*”) to the test item “*bait*” on the word-meaning recall measure or /*beit*/ and /*bi:t*/ to the same test item on the word-form recall measure were awarded with 1 and 0 point, respectively. Two experienced Vietnamese teachers of English were recruited to independently score 10% of the test papers on the word-meaning recall measure. Two experienced ESL/EFL teachers (one a native and the other a non-native speaker of English) were also invited to independently score 10% of the recordings on the word-form recall measure. Their scoring outcome was finally compared with my independent scoring. No difference was found in any between-rater comparisons of the awarded scores.

Validating the tests

This validation also uses Messickian construct validity and Rasch Model Analysis as a theoretical framework and an analytical tool, respectively. It should also be noted that these tests were straightforward in the sense that the information required to determine whether or not the participants knew the form or meaning of the potential target words was all presented in their test responses. In other words, no generalization or inference from the test scores was needed. Thus, the representativeness of the test content and the generalizability of the test scores were unnecessary in this case. In this validation, the immediate post-test scores of all participants in Study 1 were used ($N = 64$).

Content aspect

The degree to which the test items were relevant to the selected test construct was evidenced through the process of selecting and operationalizing the test construct above. As the representativeness of the test content was irrelevant in this case, the following section is only concerned with the remaining criterion - their technical quality. On the word-form recall test ($N = 28$, 14 participants from the control/comparison group and the other 14 participants from the treatment group), the item infit mean squares and the corresponding ZSTDs ranged from +0.56 to +1.30 logits and from -1.30 to +1.10 logits, respectively. The item-measure correlation coefficients ranged from .18 to .78. The internal reliability of the whole test was .73. With the only exception of Item 10 (for the word “*lure*”), which had the infit mean

square of +0.56 logit, all the other items had their infit mean squares from 0.75 to 1.30 logits. On the word-meaning recall test ($N = 36$, 18 participants from the control/comparison group and the other 18 participants from the treatment group), the item infit mean squares and the corresponding ZSTDs ranged from 0.84 to 1.29 logits and from -0.70 to 0.70 logits, respectively. The item-measure correlation coefficients were from .22 to .89. The internal reliability of the entire test was .81. All these figures fell neatly within the acceptable ranges that Baghaei and Amrahi (2011) suggested as indicators for an acceptable technical quality.

In summary, with the only exception of Item 10 in the word-form recall test, all the test items in these measures were found to be relevant to the selected construct. No problem was found for the quality of item writing, the key system or the scoring procedure. The quality of Item 10 in the word-form recall test would be reconsidered after verifying whether or not the participants had been mentally engaged with the test as expected. If participants had many unexpected responses to this item as compared to the model expectation about their vocabulary knowledge, it might be their unexpected test behaviours that had distorted the quality of this item rather than the problem of the item itself. This question will be answered in the next section.

Substantive aspect

On the word-form recall test, two participants were found to have outfit mean squares and the corresponding ZSTDs larger than 2.00 (S01 and S12, both in the control group). A closer look at their item responses showed that their answers to Item 10 were not matched with the model expectation for their vocabulary knowledge. In fact, the difficulty level of this item was at least 1 logit above the estimated values of their vocabulary knowledge, but they both got it right.

To examine whether it was their unexpected responses that had distorted the quality of this test item, the data provided by these participants were first removed from the data matrix. Rasch Model analyses were subsequently implemented again for the quality of the content aspect. There was a significant change in the item infit mean squares and the corresponding ZSTDs, which now all ranged from +0.78 to +1.26 logits and from -1.30 to +1.40 logits, respectively. This also held true for the item-measure correlation coefficients, which now were from .26 to .85. The internal reliability was now .83. Therefore, it was indeed the unexpected item responses by these participants that had distorted the quality of Item 10. Consequently, the data provided by these participants were removed from the final

analyses. On the word-meaning recall test, three other participants were also found to have outfit mean squares and the corresponding ZSDTs above 2.00. Again, their data were first excluded from the data matrix and Rasch Model analyses were carried out again on the new dataset. However, no significant improvement was found in the quality of the test items. This suggests that though these three participants had some unexpected item responses, this did not compromise the quality of the test as a whole. Given the small sample size of this research project, the data provided by these participants were all retained.

In brief, regarding the word-form recall test, two participants had unexpected responses to Item 10 as compared with the model expectation about their vocabulary knowledge. Further analyses indicated that their unexpected responses had actually distorted the statistical outcome regarding the quality of this item. Therefore, the data given by these participants were removed from the final analyses. On the word-meaning recall test, three other participants also had unexpected responses to some test items. However, this did not compromise the quality of the test as a whole. Therefore, their provided data were all retained, given the small sample size in this research project.

Structural aspect

On the word-form recall test, the total variance in the test scores explained by the test itself was 46.0%. There were five residual components in its data matrix. The proportions of the score variance that could not be explained by the test in each of these residual components ranged from 3.1% to 17.5%. On the word-meaning recall test, the total variance in the test score explained by the test was 56.2%. There were also five residual components in the data matrix. The proportions of the score variance in these components that could not be explained by the test were between 4.5% and 14.1%. All of these figures do not meet Linacre's thresholds for test unidimensionality. This suggests that there might be another meaningful construct in both of these vocabulary tests.

To further examine this issue, Stevens' (2002) guidelines were used. In all the residual components in both tests, however, there were again fewer than three, four and ten items with their absolute loadings larger than .80, .60 and .40, respectively. Therefore, the hypothesis that the current tests had a second meaningful construct was statistically rejected. In other words, the tests also met the requirement for the structural aspect of Messickian construct validity.

External aspect

As all the participants in this research project already took Nation and Beglar's VST (2007), which was successfully validated by Beglar (2009), Pearson correlation coefficients were computed to gauge the strength of association between the Rasch Model estimates of their knowledge of the potential target words and their VST scores. The results showed that the model estimates of their word-form knowledge ($N = 26$, after removing the two participants who had unexpected responses to Item 10) and their word-meaning knowledge ($N = 36$) were both significantly and positively correlated with their VST scores, with $r = .54$ ($p < .01$) and $r = .53$ ($p < .001$), respectively. Thus, both vocabulary tests met the criterion for the external aspect of construct validity.

In conclusion, the current vocabulary knowledge tests – the word-form recall test and the word-meaning recall test – were both found to include only test items directly relevant to the selected construct. No problem was found for the quality of the item writing, key system, or scoring procedure. For the word-form recall test, two participants had unexpected responses to Item 10 as compared with the model expectation for their vocabulary knowledge. This actually distorted the quality of this item and therefore their provided data were removed from the final analyses. Regarding the word-meaning recall test, three other participants also had some unexpected item responses. However, this did not compromise the quality of the test as a whole. As a result, their given data were all retained, given the small sample size in this research project. There was evidence that these two tests had only one underlying construct. In addition, the participants' scores from both tests were significantly and positively correlated with their scores on Nation and Beglar's VST (2007), which was successfully validated by Beglar (2009). In other words, both tests satisfied the requirements for Messickian construct validity. As a result, the test scores can now safely be used for further analyses.

Dependent Measures in Study 4, 5 and 7

The dependent measures in these studies also included a text comprehension test and two vocabulary knowledge tests. In what follows, the selection and operationalization of the test constructs, the design of the test items, and the validation of the test quality will be presented.

Text Comprehension Test

Designing the test

For this test, the process of selecting and operationalizing the test construct and the process of designing the test items to fit this construct were exactly the same as those of the text comprehension test for Study 1, 2, 3, and 7 described in the previous section. The only difference was that, in the present test, the content of the fifteen Yes/No questions was distributed across all three experimental videos rather than only in the longest one. Therefore, the following section is only concerned with the validation of the test quality.

Validating the test

It should be noted that this validation also used Messickian construct validity and Rasch Model as a theoretical framework and analytical tool, respectively. The text comprehension test scores of all participants in the NV condition and the NVS condition were used for this validation ($N = 64$).

Content aspect

The degree to which the test content was relevant to the selected construct was evidenced through the process of selecting and operationalizing the test construct and the process of designing the test items to fit this construct described in the previous section. In what follows, I will only report the evidence for the content representativeness and the technical quality. Figure 2 presents the estimated values of the text comprehension levels of all participants ($N = 64$) and the difficulty levels of all test items ($N = 15$) on the upper and lower chart, respectively:

Figure 2: Rasch model's Person-Item Map

As can be seen from Figure 2, the estimated values of the text comprehension levels of the participants ranged from -0.86 to $+3.13$ logits while those of the difficulty levels of the

test items ranged from -2.24 to 1.97 logits. There were six out of 68 participants (four from the NV condition and the other two from the NVS condition) whose text comprehension levels went beyond the measurement capacity of the current test. The data provided by these participants were therefore removed from the final analyses. The item strata were 3.24, which is noticeably larger than Smith's (2001) required item strata of 2 as an indicator for an acceptable level of item discrimination. The item infit mean squares and the corresponding ZSTDs were from +0.85 to +1.13 and from -1.30 to +1.70, respectively. The item-measure correlations ranged from .04 to .47. All these indices fell neatly within the acceptable ranges for an acceptable technical quality (Rodgers, 2013; Baghaei & Amrahi, 2011). The internal reliability of the whole test was .83.

It is therefore evidenced that the current test included only relevant and representative content for the selected construct. The test prompt was clearly written and included an acceptable amount of reading. No problem was found either for the key system or the scoring procedure. In other words, the current test satisfied all the criteria for the content aspect of Messickian construct validity.

Substantive aspect

There were four participants whose actual text comprehension levels were not matched with the model expectation for their L2 academic listening abilities as their outfit mean squares and the corresponding ZSTDs were all larger than 2.00 (Rodgers, 2013). Their unexpected responses to some test items might have distorted the quality of these items. To further examine this issue, the data provided by these participants were first removed from the data matrix. Rasch Model analyses were then implemented again for the quality of the content aspect. However, no significant improvement was found in this quality. This suggests that although these participants produced unexpected responses to some test items, this did not compromise the test quality as a whole. Given the small sample size of this research project, the data provided by these participants were all retained.

Alternatively, I can bring evidence for the substantive aspect by testing the initial hypotheses that I have set for the difficulty levels of the test items and the text comprehension levels of the participants in the process of test design (McLean, Kramer & Beglar, 2015). In the current test, three types of information were used to construct the test items, as follows: (a) the talk purpose and main ideas, (b) subtle, but yet important supporting ideas and key factual information, and (c) implicit meaning, logical link between ideas and speaker attitudes

towards the given topic. In his validation of a L2 academic listening test, Wagner (2002) found that his L2 test-takers scored significantly better on the items related to the purpose and main ideas of the listening materials than the items related to the supporting ideas and factual information. Their scores on these two types of test items, in turn, significantly outweighed those on the test items related to implicit meaning, logical link between ideas or speaker attitudes towards the given topics. Therefore, this finding was used to form the first hypothesis for the difficulty levels of the items in the current test.

As not all data are normally distributed, I resorted to a Kruskal Wallis test as a non-parametric counterpart of an ANOVA test for an independent-sample comparison. A significant difference was found in the test scores across the three types of test items: $X^2(2, N = 174) = 143.68$ ($p < .0001$). Conover tests were carried out for the pairwise comparisons with p values adjusted according to the family-wide error rate procedure of Holm. The results showed that the scores on the test items related to the talk purpose and main ideas were indeed significantly higher than those on the items related to the supporting ideas and factual information ($p < .0001$), which, in turn, were significantly higher than the scores on the items related to implicit meaning, logical link between ideas, and speaker attitude towards the given topic ($p < .0001$).

In addition, it was also hypothesized in the current test that the “No” items were more challenging than the “Yes” items as the participants normally focused on what was said rather than what was not said during their viewing session. Therefore, the test scores on the former were likely to be lower than those on the latter. As the dataset for this comparison was normally distributed, a t -test was used. The result indeed supported this hypothesis: $t = -5.81$ ($p < .0001$). Following the previous research findings (e.g., Bonk, 2000; Stæhr, 2009, 2008; van Zeeland & Schmitt, 2013), I also hypothesized that the larger L2 vocabulary knowledge the participants had prior to the experiment, the higher text comprehension levels they tended to obtain. A Pearson correlation coefficient was calculated for this association. Indeed, a positive coefficient of .30 ($p < .05$) was found.

Together, although four participants were not cognitively engaged with the test as expected, this did not compromise the test quality as a whole. Thus, the data provided by these participants were all retained given the small sample size in this research project. In addition, my initial hypotheses about the difficulty levels of all test items and the text

comprehension levels of all participants were all statistically supported. In other words, the current test also satisfied the requirements for the substantive aspect.

Structural aspect

The Principal Components Analysis of item residuals was carried out to examine whether the current test had one single underlying construct. The current test was found to account for only 22.3% of the variance in the test scores. There were five residual components in its data matrix. In these components, the score variance that could not be explained by the current test ranged from 7.1% to 10%. All of these figures do not meet the thresholds that Linacre (2007) suggested as indicators for item unidimensionality. Therefore, there might be a second meaningful construct in the test.

To further examine this issue, Stevens' (2002) guidelines were used. In all the residual components, however, there were consistently fewer than three, four or ten items with their absolute loadings larger than .80, .60 and .40, respectively. Therefore, the hypothesis that the current test had a second meaningful construct was statistically rejected. In other words, the test also met the requirement for the structural aspect of the construct validity.

Generalizable and external aspect

To check the generalizability of the test, the test items were first divided into two halves according to their odd and even numbering. Rasch Model analyses were then run for each half of the test to generate the estimated values for the text comprehension levels of all participants. The disattenuated correlation was 1.00, which is larger than the required disattenuated correlation of at least .90 that McLean, Kramer and Beglar (2015) proposed as an indicator for item invariance.

To validate the external aspect of the test, the estimated values of the text comprehension levels of the participants were correlated with their scores on Nation and Beglar's (2007) VST, which was successfully validated by Beglar (2009). As mentioned above, though this correlation coefficient was relatively low ($r = .30, p < .05$), it was again both significant and positive. Therefore, the test also had the external reliability.

Altogether, the current test successfully met the requirements for the content, substantive, structural, generalizable and external aspect of Messickian construct validity. Therefore, the test scores can safely be used for further analyses.

Vocabulary Knowledge Tests

Designing the tests

Again, the process of selecting and operationalizing the test construct and the process of designing the test items to fit this construct for the current tests were exactly the same as that of the vocabulary knowledge tests used for Study 1, 2, 3, and 7. The only difference was that both tests included 40 rather than 20 test items.

In what follows, I will present the validation of these tests using the immediate post-test scores of 96 participants in the control condition, the NV condition, and the NVS condition. It should be noted that this validation also used Messickian construct validity as a theoretical framework and Rasch Model as an analytical tool.

Validating the tests

Content aspect

The degree to which the test items were relevant to the selected test construct was evidenced through the process of selecting and operationalizing the test construct presented in the previous section. As the representativeness of the test content was irrelevant in this case, the following section is only concerned with the remaining criterion - their technical quality. On the word-form recall test ($N = 42$), the item infit mean squares and the corresponding ZSTDs ranged from +0.82 to +1.56 and from -1.79 to +1.81, respectively. The item-measure correlation coefficients ranged from .24 to .89. The internal reliability of the whole test was .87.

On the word-meaning recall test ($N = 54$), the item infit mean squares and the corresponding ZSTDs ranged from +0.76 to +1.12 and from -1.41 to +1.68, respectively. The item-measure correlation coefficients were from .36 to .83. The internal reliability of the whole test was .89. All of these figures fell neatly within the acceptable ranges that Baghaei and Amrahi (2011) and Rodgers (2013) suggested as indicators for an acceptable technical quality.

Therefore, the test items in both measures proved to be relevant to the selected construct. No problem was found for the quality of item writing, the key system or the scoring procedure. In other words, both measures satisfied the criteria for the content aspect of Messickian construct validity.

Substantive aspect

Four participants on the word-form recall test (one from the control group and the other three from the NV group) and three other ones on the word-meaning recall test (one from the NV group and the other two from the NVS group) were found to have the outfit mean squares and the corresponding ZSTDs larger than 2.00. This suggests that these participants might have produced unexpected responses to some test items, which, in turn, could have distorted the quality of these items.

To further investigate this issue, the data provided by these participants were first removed from the data matrix. Subsequently, Rasch Model analyses were carried out again for the quality of the content aspect. There was no significant improvement in this quality of both measures. In other words, although these participants had unexpected responses to some test items as compared to the model expectation about their vocabulary knowledge this did not compromise the test quality as a whole. Given the small sample size in this research project, the data given by these participants were all retained.

Structural aspect

The Principal Components Analysis of item residuals was implemented to examine whether each of the tests had one single underlying construct. On the word-form recall test, the total variance in the test scores explained by the test itself was 34.0%. There were five residual components in its data matrix. The proportion of the score variance that could not be explained by the test in each of these residual components ranged from 7.1% to 11.3%.

On the word-meaning recall test, the total variance in the test score explained by the test was 42.2%. There were also five residual components in the data matrix. The proportions of the score variance in these components that could not be explained by the test ranged from 5.2% to 10.3%. All of these figures do not meet Linacre's thresholds for test unidimensionality. This suggests that there might be another construct in both vocabulary measures.

To further examine this issue, Stevens' (2002) guidelines were used. In all the residual components in both tests, however, there were consistently fewer than three, four or ten items with their absolute loadings larger than .80, .60 and .40, respectively. Therefore, the hypothesis that the tests had a second meaningful construct was statistically rejected. In other words, both tests met the requirement for the structural aspect of Messickian construct validity.

External aspect

As all the participants took Nation and Beglar's VST (2007), which was successfully validated by Beglar (2009), Pearson correlation coefficients were computed to examine the relationship between their VST scores on one hand and their immediate post-test scores on the word-form and word-meaning recall measures on the other hand. The results presented a coefficient of .31 ($p < .05$) and .37 ($p < .01$), respectively. Although the coefficients were both weak, they were significant and positive. Thus, these two tests also met the criterion for the external aspect of construct validity.

In summary, the present vocabulary knowledge tests – the word-form recall test and the word-meaning recall test – also satisfied all the requirements for the content, substantive, structural, and external aspects of Messickian construct validity. As a result, the test scores were now safely used for further analyses.

Procedure of Data Collection

To collect the data for this research project, I followed the procedure presented in Table 7:

Table 7: Overall procedure of data collection

Week 1	<ul style="list-style-type: none"> - Recruiting participants; - Briefing about this research project, data collection procedure, and ethical issues; - Gathering information about recruited participants, including their latest L2 proficiency scores; - Having the recruited participants sign the consent form; - Having them complete Nation and Beglar's Vocabulary Size Test (2007)
--------	---

-
- and the pre-test for their knowledge of the potential target words;
- Selecting the target words based on the participants' pre-test scores;
 - Assigning them to the control condition, the OV condition, the RV condition, the RVS condition, the NV condition or the NVS condition.
- Week 3
- Having the participants in each learning condition carry out their corresponding viewing procedures;
 - Having them complete the corresponding text comprehension test and the corresponding immediate post-test for their knowledge of the target words;
 - Interviewing the participants recruited for Study 7 on what they think about the benefits of the output tasks for their incidental vocabulary acquisition.
- Week 5
- Having the participants complete the corresponding delayed post-test for their knowledge of the target words (except for Study 7).
- Week 9
- Giving each participant a sealed envelope of their test results;
 - Sharing useful strategies and resources for vocabulary learning in return for their involvement in this research project.
-

In Week 3, participants in each learning condition carried out their corresponding viewing procedure in their intact classes. Those in the OV condition were required to watch and note down the content of the longest experimental video (i.e., the twelve-minute one) with a view to answering content questions. After that, they had five minutes to improve their notes. Finally, they were given the 15 Yes/No content questions and allowed to consult their notes to answer these questions. The length of this classroom procedure was 27 minutes.

The participants in the RV condition followed exactly the same procedure as those in the OV condition, but did it twice. After their viewing sessions, they were provided with the same 15 Yes/No content questions and also allowed to refer to their notes while responding to those questions. This classroom procedure lasted 45 minutes.

The learners in the RVS condition, in turn, undertook the same procedure as those in the RV condition. However, instead of using the five minutes after their first viewing to elaborate and/or expand their notes, they made an oral summary of the video content and audio-recorded it with a smartphone. This classroom procedure was also 45 minutes in length.

In the NV condition, the participants were first required to watch and note down the content of the two shorter experimental videos (13 minutes) with a view to completing content questions afterwards. They were then given five minutes to make any modifications and additions to their notes. Next, they watched and took notes of the content of the longest experimental video (12 minutes), which was followed by another five minutes for their note improvement. At the end of their viewing sessions, they were provided with the other 15 Yes/No content questions and allowed to consult their notes when answering these questions. This classroom procedure also lasted 45 minutes.

Those in the NVS condition were split into two same-sized groups. Participants in each sub-group were first asked to watch and note down the content of *either* the two shorter experimental videos (13 minutes) *or* the longest experimental video (12 minutes), not only with a view to answering content questions, but also with a view to presenting an oral summary of what they had watched to a classmate from the other sub-group. Instead of being given time to amend or expand their notes, they were given 10 minutes to work in pairs and take turns presenting their summaries. They were encouraged to ask each other questions to aid their understanding of the summaries. That is why this activity is referred to as an interactional output task. All the oral summaries were audio-recorded. After giving/receiving the summaries, the learners moved on to the other video(s) that they had not yet watched themselves, but had now been told about by their peers. Finally, they were given the same 15 Yes/No content questions to answer as in the NV condition. This classroom procedure was also 45 minutes long.

The participants in the control condition followed the same procedure as those in the NV condition, but with the three non-experimental videos. Although they were not exposed to the target words in their input materials, they still took the same vocabulary knowledge tests as the participants in the treatment conditions. Vocabulary gain (if any) attested in this condition would be set as a baseline to compare with vocabulary gains found in the treatment conditions.

In Study 7, four participants followed the same procedure as those in the RVS condition while the other four undertook the same procedure as those in the NVS condition. At their treatment completion, each of them was involved in a semi-structured one-to-one interview with me in their first language – Vietnamese – about the cognitive and metacognitive processes that they had used to deal with the output tasks and how these processes had affected their text comprehension and incidental vocabulary uptake.

With the only exception of the OV condition where the learners only viewed the longest experimental video and did it only once, the time invested in the other learning conditions was kept constant at 45 minutes. The learners were all informed beforehand about their corresponding treatment procedures and the follow-up content questions. However, the presence of the two post-tests on the target words was not announced as this research project examines incidental rather than intentional vocabulary learning.

It should be also noted that, unlike most previous research in this area, the present research project used an oral rather than a written summary activity as a pushed output task. There were three reasons for this choice. First, to summarize the input content, participants in this research project may need to use some newly met words from their preceding listening stage. In such a case, it is invalid to ask them to recall the written form of these words as they were not exposed to this form in their listening materials. Second, it is easier to prompt learner-learner interaction (as in the case of Study 5) by means of an oral summary activity than a written summary activity. Finally, a written summary activity is generally more time-consuming than an oral summary activity. This, in turn, might have negatively affected my control of time on task across all treatment conditions.

Data Analysis

The purpose of the vocabulary pre-test was to establish a set of the target words which were unknown to the participants prior to the experiment. Following the studies by R. Ellis and He (1999) and R. Ellis et al. (1994), for example, the words that were already known by more than 10% of the participants were removed from the final analyses. After this removal, in the case of Study 1, the number of the target words was 13 on the word-form recall measure and 18 on the word-meaning recall measure. These numbers in the case of Study 2 and 3 were 17 and 18 and in the case of Study 4 and 5 they were 25 and 29, respectively. In Study 7, these figures widely varied from one participant to another and will be reported in detail in the following chapter.

For the descriptive statistics on the vocabulary measures, I shall report the sample size (N), the mean score (M), and the corresponding standard deviation (SD). Henceforth, I will use the words “*gain*” and “*uptake*” interchangeably as umbrella terms for any incidental L2 vocabulary learning evidenced in the post-tests. In addition, the terms “*acquisition*” and “*retention*” will be used specifically to refer to the learning outcome as evidenced in the immediate post-test and the delayed post-test, respectively. The size of vocabulary acquisition is calculated by subtracting the pre-test scores from the immediate post-test scores while the size of vocabulary retention is determined by subtracting the pre-test scores from the delayed post-test scores. The absolute rather than relative vocabulary gain is used in this research project in order not to inflate the true effect of each learning condition (*cf.* Rodgers, 2013; Webb & Chang, 2012, 2014, 2015).

Cohen’s effect size d will be used to gauge the effect size of each treatment condition on vocabulary acquisition and retention. This effect size is computed by the following formula: $d = (M_1 - M_2) / s_{\text{pooled}}$ ($s_{\text{pooled}} = \sqrt{[(s_1^2 + s_2^2)/2]}$, $M_{1,2}$ = mean score, $S_{1,2}$ = standard deviation). According to Cohen (1977), a d value of .20, .50, and .80 is normally used as an indicator for a small, medium and large effect size, respectively.

As not all test data are normally distributed, I resort to Friedman tests as the non-parametric counterpart of the correlated-samples ANOVA tests to examine the difference in the vocabulary test scores across the three testing times – the pre-test, the immediate post-test, and the delayed post-test – per learning condition. A two-tailed p value of .05 is set as a threshold for significance in these tests. As Friedman tests are not accompanied by any post-hoc tests for pairwise comparisons, Wilcoxon Signed-Rank tests will be used for this purpose. Two-tailed p values of .05 (without the Bonferroni post-hoc correction) and .017 (with the Bonferroni post-hoc correction) are set as two different thresholds for significance in these tests.

To examine the difference in the size of vocabulary gain between the control condition and the treatment condition, Mann-Whitney tests will be used as the non-parametric counterpart of the t -tests for independent sample comparisons. A two-tailed p value of .05 is set as a threshold for significance in these tests.

In addition, my report of research results will also include Pearson correlation computation and Multiple regression analysis to examine the associations between vocabulary gain on the one hand and learners’ text comprehension, their prior L2 proficiency

and vocabulary knowledge, the frequency of word occurrences in the input materials, and the usefulness of verbal and pictorial clues in the input materials for word-meaning inference on the other hand. A two-tailed p value of .05 is also set as a threshold for significance in these calculations.

For Study 6, the Cohen's effect size on vocabulary gain found for each learning condition from Study 1 to Study 5 will be incorporated into the meta-analysis program in the Exploratory Software for Confidence Intervals (ESCI) developed by Cumming (2012) to rank these conditions according to their true effect sizes. Multiple regressions will also be used to identify the factor or factors that best explain(s) the difference in these effect sizes (if any).

Regarding the qualitative data, I myself first listened to the audio-recordings of all the interviews several times, noted down all cognitive and metacognitive processes that the participants had used to deal with the summary tasks and how these processes had affected their text comprehension and incidental vocabulary acquisition, and then built up a checklist of the coding results, together with the corresponding evidence in the form of direct quotations of what the participants actually said. Two experienced Vietnamese teachers of EFL/ESL were invited to listen to all recordings and use the checklist above to put a tick or a cross next to each process, effect and/or evidence. In case they had a different opinion upon a particular process, effect, and evidence or recognized any new process, effect and evidence, they were required to note these down on a blank sheet. When they completed their checklist, I had an in-depth discussion with them about any differences in our coding results. After this discussion, we together constructed a final list of the processes, effects, and evidence that we all agreed upon (see Appendix K). These processes, effects, and evidence will be reported in detail in Study 7 in the following chapter - RESULTS.

CHAPTER 3: RESULTS

In this chapter, I shall report the results from the five quantitative studies in this research project from Section 1 to Section 5. Each study report will first give the results on the size of vocabulary gain made from each treatment condition. This vocabulary gain is then compared with the baseline gain made in the control condition, where the learners did not view the experimental video(s) and therefore were not exposed to the target words. Gain (if any) between the pre-test and the post-tests by the control group can therefore only be attributed to test-taking, chance encounters with the target words between the testing times, or sheer coincidence. If such a gain is attested in the control condition, then this needs to be taken into account when interpreting the gain observed in the treatment condition, because this would mean that not all of the latter are necessarily the result of the treatment as such. The associations between the vocabulary gain on the one hand and the learners' text comprehension, their prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input materials, and the usefulness of verbal and pictorial clues in these materials for word-meaning inference on the other hand will also be covered in these sections. In Section 6, a series of statistical comparisons and factorial analyses will be carried out to examine the difference in the effect sizes on vocabulary gain across all treatment conditions, and then look for the explanatory factors for this difference (if any). Subsequently, Section 7 will provide the results from the qualitative study, regarding the cognitive and metacognitive processes that the learners used to perform the monologue summary activity in the RVS condition and the interactional summary activity in the NVS condition and how these processes affected their text comprehension and incidental vocabulary acquisition. In the final section, the results of the whole research project will be summarized in a comprehensive table.

Study 1: Viewing a TED Talks Video Once (OV)

The aim of this study was twofold. First, it measured the size of vocabulary gain from a single viewing of a TED Talks video (i.e., the OV condition). Second, it examined the association between this vocabulary gain and the learners' text comprehension. Third, the relationships between this vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input material, and the usefulness of verbal and pictorial clues in this material for word-meaning

inference on the other hand were also explored in this study. The research questions investigated in the present study were as follows:

- (a) *To what extent do L2 learners incidentally pick up the form and meaning of novel words from viewing a TED Talks video once?*
- (b) *Is this vocabulary gain associated with the learners' text comprehension?*
- (c) *Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference?*

Word-form Gain

Table 8 presents the descriptive statistics for the scores that the learners in the control group and the treatment group obtained on the word-form recall measure. Specifically, it includes the sample sizes (*N*), the mean scores (*M*), and the corresponding standard deviations (*SD*) as gauged on the pre-test, the immediate post-test, and the delayed post-test. This table also includes the acquisition scores (as calculated by subtracting the pre-test scores from the immediate post-test scores), the retention scores (as computed by subtracting the pre-test scores from the delayed post-test scores), and the corresponding standard deviations. In the case of the control group, the data were only provided by 12 out of the 14 learners because the test validation in the Methodology Chapter revealed that the vocabulary knowledge of two learners was not effectively measured by the current test and therefore their test scores should be removed from the final analyses. After excluding the target words that were already known to more than 10% of the learners, the data on this measure were now related to only 13 out of the 20 target words.

Table 8: Word-form gain in OV – 13 target words (max. = 13 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	12	0.07	0.27	14	0.00	0.00
Post-test	12	0.07	0.27	14	0.57	0.51
Delayed post-test	12	0.07	0.27	14	0.57	0.51
Acquisition	12	0.00	0.00	14	0.57	0.51
Retention	12	0.00	0.00	14	0.57	0.51

To examine the difference in the learners' prior knowledge of the form of the target words between the groups, a Mann Whitney test was carried out with their pre-test scores. However, no difference was found: $U = 91$ ($p = .76$). Therefore, it is now safer to attribute any difference in the size of their word-form gains to the effects of the learning conditions.

It is clear from Table 8 that the learners in the treatment condition made a very limited word-form gain, of only 0.57 target words, on both immediate and delayed post-tests. To measure the effect size of this condition on word-form gain, Cohen's d values were calculated. The results, however, showed that this condition in fact produced a large effect size of 1.52 on both word-form acquisition and word-form retention despite the small number of the target words that the learners actually recalled the form of in the post-tests.

Friedman tests were carried out for each learning condition to gauge the difference in the learners' vocabulary test scores across the three testing times – the pre-test, the immediate post-test, and the delayed post-test. No difference was found for the control group: $X^2(2, N = 12) = 0.00$ ($p = 1.000$). However, a significant difference was found in the case of the treatment group: $X^2(2, N = 14) = 6.86$ ($p < .05$). Wilcoxon Signed-Rank tests were subsequently implemented for the pairwise comparisons in this group. The results showed that their immediate and delayed post-test scores were identical and both significantly surpassed their pre-test scores with the same Z value of 2.52 ($p < .017$).

To compare the size of word-form gains between the groups, two Mann Whitney tests were implemented with their word-form acquisition and retention scores. As expected, the learners in the treatment group indeed achieved significantly higher word-form acquisition and retention scores than those in the control group with the same U value of 154 ($p < .05$).

Word-meaning Gain

Table 9 presents the descriptive statistics for the scores that the learners in these groups achieved on the word-meaning recall measure. After removing the target words that were already known to more than 10% of the learners, the data on this measure were now related to only 18 out of the 20 target words.

Table 9: Word-meaning gain in OV – 18 target words (max. = 18 points)

Scores	Control condition			Treatment condition		
	N	M	SD	N	M	SD

Pre-test	18	0.33	0.49	18	0.22	0.43
Post-test	18	0.33	0.49	18	0.50	0.52
Delayed post-test	18	0.33	0.49	18	0.44	0.51
Acquisition	18	0.00	0.00	18	0.28	0.46
Retention	18	0.00	0.00	18	0.22	0.43

To examine whether the learners' pre-test scores were significantly different between the groups, a Mann Whitney test was carried out. However, no difference was found: $U = 153$ ($p = .79$). This suggests that the learners in the groups had a roughly equal amount of knowledge about the meaning of the target words prior this experiment.

As can be seen from Table 9, the number of target words that the learners in the treatment condition recalled the meaning of in the two post-tests was very small. Cohen's d values were computed to gauge the effect size of this condition on word-meaning gain. Despite the small gain in real terms, this condition was found to create a large effect size of .86 and a medium effect size of .72 on word-meaning acquisition and retention, respectively.

Friedman tests were run for each learning condition to measure the difference in the learners' vocabulary test scores across the three testing times. However, neither was a difference found in the control condition ($X^2(2, N = 18) = 0.00, p = 1.00$) nor the treatment condition ($X^2(2, N = 18) = 1.75, p = .42$).

To compare the size of word-meaning gains between the groups, two Mann Whitney tests were implemented for their word-meaning acquisition and retention scores. The results showed that the word-meaning acquisition scores in the treatment condition were noticeably higher than those in the control condition, but this difference fell short of the required significance level: $U = 198$ ($p = .26$). This also held true for the case of their word-meaning retention scores: $U = 189$ ($p = .40$).

Association between Vocabulary Gain and Text Comprehension

The data on the text comprehension measure were only provided by 31 out of the 32 learners as the test validation (see the Methodology Chapter) showed that the text comprehension level of one learner was not effectively measured by the current test and therefore the test scores obtained by this learner should be removed from the final analyses. For the remaining learners, they were found to obtain a mean score of 5.13 (SD 1.43) out of the maximum 15 points (34.20%). Their test scores ranged from 3.00 (20%) to 8.00 (53.33%)

points. It should be noted that there was only one learner with the score of 8 points. In other words, the learners in this study achieved a relatively limited degree of text comprehension after viewing the video once.

The learners' modest text comprehension might explain their small vocabulary gain. To examine the plausibility of this account, Pearson correlation coefficients were calculated to gauge the direction, strength, and significance of the association between the text comprehension scores on the one hand and the vocabulary acquisition and retention scores on the other hand. On the word-form recall test, these coefficients were exactly the same: $r = .06$ ($p = .84$). On the word-meaning recall test, these coefficients were $r = .04$ ($p = .88$) and $r = .06$ ($p = .81$), respectively. Although these coefficients were all positive, they were consistently weak and non-significant.

Associations between Vocabulary Gain and other Factors of Interest

Pearson correlation coefficients were calculated to gauge the direction, strength, and significance of the associations between the vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge on the other hand. In these cases, the size of vocabulary gain was computed from the gain score per person. The associations between the vocabulary gain on the one hand and the frequency of word occurrences and the usefulness of verbal and pictorial clues in the input material for word-meaning inference on the other hand were also examined using Pearson correlation coefficients. However, the size of lexical uptake in such cases was summed up from the gain score per item. The results from the above analyses are summarized in Table 10. As can be seen from the table, all coefficients were again positive, but all weak and non-significant:

Table 10: Association between vocabulary gain in OV and factors of interest

		L2 proficiency	Vocabulary knowledge	Word occurrences	Verbal clues	Pictorial clues
Form	Acquisition	.04	.13	.13	.22	.06
recall	Retention	.14	.13	.13	.22	.06
Meaning	Acquisition	.22	.30	.05	.05	.25
recall	Retention	.21	.31	.09	.12	.07

None reaches the required significance level ($p = .05$)

Summary

The report above indicates that the L2 learners in this study were able to make a statistically significant gain in terms of word form, but not word meaning, from viewing the TED Talks video once. In real terms, this gain was very modest, however. On average, these learners picked up only 0.5 target word from this learning condition. The association between this vocabulary gain and the learners' text comprehension was positive, but far from linear. This also holds true for the relationships between this vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference on the other hand.

Study 2: Adding another Viewing of the Same Video (RV)

This study gauged the size of vocabulary gain from viewing the same TED Talks video as in Study 1, but doing it twice (i.e., the RV condition). This study also aimed to examine the association between this vocabulary gain and the learners' text comprehension. It also took another look at the correlations between this vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input material, and the usefulness of verbal and pictorial clues in the input material for word-meaning inference on the other hand. This study addressed the following research questions:

- (d) To what extent do L2 learners incidentally pick up the form and meaning of novel words from viewing a TED Talks video twice?*
- (e) Is this vocabulary gain associated with the learners' text comprehension?*
- (f) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference?*

Word-form Gain

Table 11 presents the descriptive statistics for the scores that the learners in the control group and the treatment group obtained on the word-form recall measure. After the target words that were already known to more than 10% of the learners were removed, the data on this measure were now related to only 17 out of the 20 target words.

Table 11: Word-form gain in RV – 17 target words (max. = 17 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	12	0.21	0.43	14	0.29	0.47
Post-test	12	0.21	0.43	14	1.14	0.95
Delayed post-test	12	0.21	0.43	14	1.07	0.91
Acquisition	12	0.00	0.00	14	0.85	0.69
Retention	12	0.00	0.00	14	0.78	0.57

A Mann Whitney test was implemented to compare the learners' pre-test scores between the groups. No difference was found: $U = 105$ ($p = .76$). This ensures that the learners in the groups had a roughly equal starting point regarding their knowledge of the form of the target words before they participated in this experiment.

The descriptive statistics in Table 11 show that the size of word-form gain in this condition proved to be larger than that in the OV condition as reported in Study 1. However, this gain still remained small as the learners recalled the form of only one target word on average after the intervention. To examine the effect size of this condition on vocabulary gain, Cohen's d values were calculated. The results showed that this condition nonetheless had a large effect of 1.58 and 1.56 on word-form acquisition and retention, respectively.

Friedman tests were carried out to compare the learners' vocabulary test scores for each condition across the three testing times. No difference was found for the control condition: $X^2(2, N = 12) = 0.00$ ($p = 1.00$). However, the scores obtained by the learners in the treatment condition significantly increased from the pre-test to the post-tests: $X^2(2, N = 14) = 8.13$ ($p < .05$). Three Wilcoxon Signed-Rank tests were run for the pairwise comparisons in this condition and indicated that the scores in the two post-tests were almost the same and they were both significantly higher than the pre-test scores with the same Z value of 2.67 ($p < .017$).

To compare the size of word-form gains between the groups, two Mann Whitney tests were conducted with their acquisition and retention scores. The learners in the treatment group were indeed found to make significantly larger word-form acquisition and retention than those in the control group with the same U value of 168 ($p < .01$).

Word-meaning Gain

Table 12 presents the descriptive statistics for the scores that the learners in the two groups obtained on the word-meaning recall measure. Again, the removal of the target words that were already known to more than 10% of the learners reduced the original number of the target words on this measure from 20 to 18.

Table 12: Word-meaning gain in RV – 18 target words (max. = 18 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	18	0.33	0.49	18	0.28	0.58
Post-test	18	0.33	0.49	18	1.06	1.06
Delayed post-test	18	0.33	0.49	18	0.89	1.02
Acquisition	18	0.00	0.00	18	0.78	0.65
Retention	18	0.00	0.00	18	0.61	0.76

To ensure that there was no significant difference in the learners' prior knowledge of the meaning of the target words between the groups, a Mann Whitney test was run. The result indeed showed that the learners in these groups had a similar amount of knowledge about the meaning of the lexical items before they were involved in this experiment: $U = 139$ ($p = .47$).

Cohen's d values were calculated to gauge the effect size of the treatment condition on word-meaning gain. This learning condition was found to create a large effect size of 1.70 and 1.14 on word-meaning acquisition and retention, respectively. However, it should be again acknowledged that, in real terms, the number of target words whose meaning was successfully recalled in the two post-tests was still limited, as shown in Table 12.

To examine whether the learners' scores significantly increased from the pre-test to the post-tests, Friedman tests were carried out for each learning condition. In the control condition, these scores remained constant: $X^2(2, N = 18) = 0.00$ ($p = 1.00$). However, a significant score improvement was found for the case of the treatment condition: $X^2(2, N = 18) = 8.55$ ($p < .05$). Three Wilcoxon Signed-Rank tests were subsequently implemented for the pairwise comparisons in this condition and indicated that the scores in the immediate post-test and the delayed post-test did not significantly differ ($Z = .80$, $p = .42$), but both were significantly higher than the pre-test scores with $Z = 3.06$ ($p < .017$) and $Z = 2.52$ ($p < .017$), respectively.

Mann Whitney tests were conducted to compare the size of word-meaning gain between the conditions. The results showed that the treatment condition led to significantly larger word-meaning acquisition ($U = 263$, $p < .01$) and retention ($U = 243$, $p < .05$) than the control condition.

Association between Vocabulary Gain and Text Comprehension

On the text comprehension measure, the learners in the treatment condition ($N = 32$) obtained a mean score of 7.25 (SD 1.24) out of the maximum score of 15 points (48.33%). Their test scores ranged from 3 (20%) to 12 (80%) points. Pearson correlation coefficients were computed to gauge the direction, strength, and significance of the association between vocabulary gain and text comprehension.

On the word-form recall measure, the coefficients between the text comprehension scores on the one hand and the acquisition and retention scores on the other hand were .54 ($p < .05$) and .46 ($p = .10$), respectively. These coefficients in the case of the word-meaning recall measure were .49 ($p < .05$) and .58 ($p < .05$). Thus, the more successfully the learners understood and remembered the video content, the larger the vocabulary gain they tended to make.

Associations between Vocabulary Gain and other Factors of Interest

Table 13 presents the Pearson correlation coefficients between the vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input material, and the usefulness of verbal and pictorial clues in the input material for word-meaning inference on the other hand:

Table 13: Association between vocabulary gain in RV and factors of interest

		L2 proficiency	Vocabulary knowledge	Word occurrences	Verbal clues	Pictorial clues
Form	Acquisition	.31	.03	.14	.11	.23
recall	Retention	.34	.09	.14	.11	.23
Meaning	Acquisition	.35	.34	.51*	.41	.28
recall	Retention	.41	.32	.53*	.45	.23

* $p < .05$

As can be seen from this table, only the frequency of word occurrences in the input material was significantly correlated with the size of vocabulary gain. Specifically, the more

often a target word occurred in the video, the more likely it was to be learnt. It should be also noted that the coefficients between the size of vocabulary gain on the one hand and the other factors of interest were also positive, but all fell short of the required significance level.

Summary

It is clear from the report above that having L2 learners view the same TED Talks video twice led to a small but nonetheless statistically significant vocabulary gain. On average, the learners in this study picked up one target word after this viewing procedure. This vocabulary gain was found to be positively correlated with the learners' text comprehension. A positive correlation was also found between the frequency of word occurrences in the input material and the size of lexical uptake, but only in the case of the word-meaning gain.

Study 3: Adding an Output Task between these Two Viewing Times (RVS)

This study measured the size of vocabulary gain from a learning condition where L2 learners viewed the same TED Talks video twice as in Study 2, but with an output task in the form of an oral summary activity inserted in between (i.e., the RVS condition). In addition, it also investigated the correlation between this vocabulary gain and the learners' text comprehension. Like the studies above, the present study also looked at the associations between this vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input material, and the usefulness of verbal and pictorial clues in the input material for word-meaning inference on the other hand. The research questions examined under this study were as follows:

- (g) To what extent do L2 learners incidentally pick up the form and meaning of novel words from viewing a TED Talks video, making an oral summary of the content, and then viewing this video again?*
- (h) Is this vocabulary gain associated with the learners' text comprehension?*
- (i) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the video, and the usefulness of verbal and pictorial clues in this video for word-meaning inference?*

Word-form Gain

The descriptive statistics for the scores that the learners in the control group and the treatment group in this study obtained on the word-form recall measure are presented in Table 14. It should be noted that three target words were already known to more than 10% of the learners and therefore removed from the final analyses.

Table 14: Word-form gain in RVS – 17 target words (max. = 17 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	12	0.21	0.43	14	0.29	0.61
Post-test	12	0.21	0.43	14	2.21	2.12
Delayed post-test	12	0.21	0.43	14	2.29	2.09
Acquisition	12	0.00	0.00	14	1.92	1.64
Retention	12	0.00	0.00	14	2.00	1.62

A Mann Whitney test was implemented to examine the difference in the learners' pre-test scores between the groups. No difference was found: $U = 100$ ($p = .96$). This suggests that the learners in two groups had a nearly identical amount of knowledge of the form of the target words prior to this experiment.

The average number of target words that the learners in the treatment condition could successfully recall the form of in the post-tests was around two. This figure was noticeably higher than that in the RV condition as reported in Study 2. To gauge the effect size of this condition on word-form gain, Cohen's d values were computed. The results showed that this condition produced a large effect size of 1.60 and 1.68 on word-form acquisition and retention, respectively.

Friedman tests were carried out to compare the learners' test scores in each learning condition across the three testing times. The test scores obtained by the learners in the control condition remained exactly the same from the pre-test to the post-tests: $X^2(2, N = 12) = 0.00$ ($p = 1.00$). Meanwhile, the test scores by those in the treatment condition significantly increased from the pre-test to the post-tests: $X^2(2, N = 14) = 15.46$ ($p < .001$). The results from the Wilcoxon Signed-Rank tests revealed that there was no significant difference in the two post-test scores, but these post-test scores significantly surpassed the pre-test scores with the same Z value of 3.30 ($p < .017$).

Mann Whitney tests were run to compare the size of word-form gains between the conditions. The results showed that the treatment condition resulted in significantly better word-form acquisition and retention than the control condition with the same $U = 182$ ($p < .001$).

Word-meaning Gain

Table 15 gives the scores obtained by the two groups on the word-meaning recall measure. After excluding the target words that were already known to more than 10% of the learners, the data on this measure were now related to only 18 out of the 20 target words.

Table 15: Word-meaning gain in RVS – 18 target words (max. = 18 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	18	0.39	0.50	18	0.17	0.38
Post-test	18	0.44	0.62	18	1.94	1.39
Delayed post-test	18	0.39	0.50	18	1.89	1.23
Acquisition	18	0.06	0.24	18	1.79	1.22
Retention	18	0.00	0.00	18	1.72	1.07

To examine the difference in the learners' pre-test scores between the groups, a Mann Whitney test was implemented. No difference was found: $U = 144$ ($p = .58$). It is nevertheless worth noting that the learners in the control group had a slightly larger amount of knowledge about the meaning of the target words than those in the treatment condition.

The descriptive statistics show that, on average, the learners in the treatment condition were able to recall the meaning of two target words in the post-tests. Cohen's d values were calculated to gauge the effect size of this condition on word-meaning gain. The results showed that this condition created a large effect size of 1.97 and 2.27 on word-meaning acquisition and retention, respectively.

To measure the difference in the learners' scores across the three testing times, Friedman tests were carried out. While no significant difference was found in the control condition ($X^2(2, N = 18) = 0.08, p = .96$), a significant difference was found in the case of the treatment condition ($X^2(2, N = 18) = 24.11, p < .0001$). The results from the Wilcoxon Signed-Rank tests revealed that the immediate and delayed post-test scores in this condition

were nearly the same and both significantly outweighed the pre-test scores with the same Z value of 3.82 ($p < .017$).

Again, Mann Whitney tests were used to compare the size of word-meaning gains between the groups. The learners in the treatment condition were found to obtain significantly larger word-meaning acquisition ($U = 311, p < .0001$) and retention ($U = 315, p < .0001$) than those in the control condition.

Association between Vocabulary Gain and Text Comprehension

On the text comprehension measure, the learners in the treatment group ($N = 32$) achieved a mean score of 9.32 ($SD 1.18$) out of the maximum score of 15 points (62.13%). Their test scores ranged between 6 (40%) and 13 (86.57%) points. To examine the direction, strength, and significance of the association between vocabulary gain and text comprehension, Pearson correlation coefficients were computed.

On the word-form recall test, the coefficients between the text comprehension scores on the one hand and the word-form acquisition and retention scores on the other hand were .62 ($p < .05$) and .59 ($p < .05$), respectively. In the case of the word-meaning recall test, these coefficients were the same: $r = .50$ ($p < .05$). Therefore, the more successfully the learners took in the video content, the larger the amount of vocabulary knowledge they were likely to gain.

Associations between Vocabulary Gain and other Factors of Interest

Pearson correlation coefficients were also calculated to gauge the direction, strength, and significance of the associations between the vocabulary gain in the treatment condition on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input material, and the usefulness of verbal and pictorial clues in this material for word-meaning inference on the other hand. The results are summarized in Table 16 below:

Table 16: Associations between vocabulary gain in RVS and factors of interest

		L2	Vocabulary	Word	Verbal	Pictorial
		proficiency	knowledge	occurrences	clues	clues
Form	Acquisition	.38	.02	.57*	.56*	.44
recall	Retention	.41	.05	.57*	.56*	.44

Meaning	Acquisition	.35	.24	.47*	.49*	.12
recall	Retention	.27	.21	.59**	.58*	.18

* $p < .05$; ** $p < .01$

This table shows that the size of vocabulary gain in this learning condition was positively correlated with the frequency of word occurrences in the video and the usefulness of verbal clues in this video for word-meaning inference. It should be also noted that the associations between the vocabulary gain and the other factors of interest were also positive, but all fell short of the required significance level.

Summary

The results from this study indicate that having L2 learners view the video, summarize the content, and then view this video again indeed brought about a statistically significant vocabulary gain. The learners, on average, picked up two target words from this viewing procedure. This vocabulary gain was found to be positively correlated with their text comprehension. The size of vocabulary gain in this study might also be dependent on the frequency of word occurrences in the video and the usefulness of verbal clues in this video for word-meaning inference.

Study 4: Adding Two other Videos of Similar Topics (NV)

First, this study aimed to gauge the size of vocabulary gain from viewing the same TED Talks video as in the studies above, but being preceded by two other videos of similar topics (i.e., the NV condition). In fact, this vocabulary gain was calculated from two different datasets. One focused on the gain obtained from the 20 target words in the longest video while the other was related to the lexical uptake made out of the 40 target words that occur in all three videos. Second, it also examined the relationship between the vocabulary gain from all three videos and the learners' text comprehension. The associations between this vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input materials, and the usefulness of verbal and pictorial clues in the input materials for word-meaning inference on the other hand were also looked at. This study addressed the following research questions:

- (j) *To what extent do L2 learners incidentally pick up the form and meaning of novel words from the longest video and from all three videos in this intervention?*

- (k) *Is the vocabulary gain from all three videos associated with the learners' text comprehension?*
- (l) *Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the videos, and the usefulness of verbal and pictorial clues in these videos for word-meaning inference?*

Word-form Gain from the Longest Video

Table 17 presents the descriptive statistics for the scores that the learners in the control group and the treatment group obtained on the word-form recall measure for the 20 target words in the longest video. After the target words that were already known to more than 10% of the learners were removed, the data in this table were only related to 14 out of the 20 target words.

Table 17: Word-form gain from the longest video – 14 target words (max. = 14 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	14	0.07	0.27	14	0.07	0.27
Post-test	14	0.07	0.27	14	0.64	0.50
Delayed post-test	14	0.07	0.27	14	0.57	0.51
Acquisition	14	0.00	0.00	14	0.57	0.51
Retention	14	0.00	0.00	14	0.50	0.52

To ensure that there was no significant difference in the learners' prior knowledge about the form of the target words between these two groups, a Mann Whitney test was carried out with their pre-test scores. No difference was found: $U = 98$ ($p = .98$).

Cohen's d values were computed to gauge the effect sizes of the treatment condition on the word-form acquisition and retention. Although this condition brought about a very limited vocabulary gain of merely around 0.50 target word on this measure, the effect sizes it created on word-form acquisition and retention were both large, with $d = 1.58$ and 1.36 , respectively.

Friedman tests were implemented to compare the scores obtained by the learners in each condition across the three testing times. In the case of the control condition, these scores were found to remain the same: $X^2(2, N = 14) = 0.00$ ($p = 1.00$). However, the learners in the

treatment condition significantly increased their scores from the pre-test to the post-tests: $X^2(2, N = 14) = 6.11$ ($p < .05$). Wilcoxon Signed Rank tests were subsequently run for the pairwise comparisons in this condition. The results showed that the immediate post-test scores and the delayed post-test scores were virtually the same and these scores were both significantly higher than the pre-test scores with $Z = 2.52$ ($p < .05$) and $Z = 2.37$ ($p < .05$), respectively.

To compare the size of the word-form gains between these two conditions, two Mann Whitney tests were conducted with their acquisition and retention scores. The treatment condition indeed led to significantly larger word-form acquisition and retention than the control condition with a U value of 154 ($p < .05$) and of 147 ($p < .05$), respectively.

To sum up, though the number of target words that the learners in the treatment condition picked up on this measure was relatively limited, this condition indeed created a large effect size on both word-form acquisition and retention.

Word-meaning Gain from the Longest Video

The scores that the learners in the two groups achieved on the word-meaning measure are summarized in Table 18. The removal of the target words which were already known to more than 10% of the learners reduced the number of the target words on this measure from 20 to 15.

Table 18: Word-meaning gain from the longest video – 15 target words (max. = 15 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	18	0.22	0.43	18	0.17	0.38
Post-test	18	0.22	0.43	18	0.72	0.92
Delayed post-test	18	0.22	0.43	18	0.72	0.92
Acquisition	18	0.00	0.00	18	0.55	0.78
Retention	18	0.00	0.00	18	0.55	0.78

To compare the learners' prior knowledge about the meaning of the target words between the two groups, a Mann Whitney test was implemented with their pre-test scores. The result indicated that the learners in these two groups had a roughly equal amount of this knowledge before they took part in this experiment: $U = 171$ ($p = .79$).

Table 19 indicates that the numbers of the target words whose correct meaning was recalled by the learners in the treatment condition in the two post-tests were both limited. To examine the effect sizes that this condition had on the word-meaning acquisition and retention, Cohen's d values were calculated. These effect sizes were found to be large with the same $d = 1.00$ for both the word-meaning acquisition and retention.

Friedman tests were carried out to compare the test scores that the learners in each condition obtained across the three testing times. Again, these scores kept unchanged in the case of the control condition: $X^2(2, N = 18) = 0.00$ ($p = 1.00$). However, the scores that the learners in the treatment condition obtained on this measure significantly improved from the pre-test to the post-tests: $X^2(2, N = 18) = 2.52$ ($p < .05$). Their scores were exactly the same between the immediate post-test and the delayed post-test.

To examine whether the treatment condition brought about a larger vocabulary gain than the control condition, two Mann Whitney tests were carried out with their word-meaning acquisition and retention scores. As expected, the learners in the treatment condition indeed successfully recalled the meaning of more target words than those in the control condition in both immediate and delayed post-tests with the same U value of 234 ($p < .05$).

The report above shows that though the numbers of the target words whose meaning was incidentally acquired and retained by the learners in the treatment condition were small, this condition, in fact, produced a large effect size on both word-meaning acquisition and retention.

Word-form Gain from all Three Videos

The descriptive statistics for the scores obtained by the learners in the control group and the treatment group on the word-form recall measure are given in Table 19. It should be noted that the data on this measure were only related to 25 out of the 40 target words after removing 15 words already known by more than 10% of the learners.

Table 19: Word-form gain from three videos – 25 target words (max. = 25 points)

Scores	Control condition			Treatment condition		
	N	M	SD	N	M	SD
Pre-test	14	0.36	0.50	14	0.36	0.50
Post-test	14	0.50	0.65	14	1.14	0.86

Delayed post-test	14	0.43	0.51	14	1.14	0.86
Acquisition	14	0.14	0.36	14	0.79	0.70
Retention	14	0.07	0.26	14	0.79	0.70

A Mann Whitney test was carried out to compare the learners' prior knowledge of the form of the target words between the groups. However, no difference was found: $U = 98$ ($p = .98$). This means that the learners in the two groups started this experiment with a similar amount of knowledge regarding the form of the target words.

Table 19 shows that, on average, the learners in the treatment condition picked up one target word after their viewing procedure. To gauge the effect size of this condition on word-form gain, Cohen's d values were computed. This condition was found to produce a large effect size of 1.17 on word-form acquisition and of 1.36 on word-form retention.

To examine whether the learners' vocabulary test scores significantly grew from the pre-test to the post-tests, Friedman tests were implemented for each learning condition. In the control condition, no significant score change was found: $X^2(2, N = 14) = 0.32$ ($p = .85$). Meanwhile, a significant score increase was found for the case of the treatment condition: $X^2(2, N = 14) = 8.68$ ($p < .05$). Three Wilcoxon Signed-Rank tests were then implemented for the pairwise comparisons in this condition. The results indicated that the scores in the two post-tests were identical and both were markedly higher than the pre-test scores. However, this difference only reached significance when the Bonferroni post-hoc correction was not applied: $Z = 2.67$ ($.017 < p < .05$).

Two Mann Whitney tests were conducted to compare the word-form acquisition and retention scores between the groups. As hypothesized, the learners in the treatment group achieved significantly higher word-form acquisition and retention scores than those in the control group with the same U value of 161 ($p < .01$).

Word-meaning Gain from all Three Videos

Table 20 sums up the scores obtained by the learners in the two groups on the word-meaning recall measure. After removing the target words that were already known to more than 10% of the learners, the data on this measure were now related to only 29 out of the 40 target words.

Table 20: Word-meaning gain from three videos – 29 target words (max. = 29 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	18	0.50	0.71	18	0.50	0.71
Post-test	18	0.50	0.71	18	1.17	0.99
Delayed post-test	18	0.50	0.71	18	1.06	0.80
Acquisition	18	0.00	0.00	18	0.67	0.91
Retention	18	0.00	0.00	18	0.56	0.71

To ascertain that these groups were not different in terms of their prior knowledge about the meaning of the target words, a Mann Whitney test was conducted with their pre-test scores. Indeed, no difference was found: $U = 144$ ($p = .58$).

The results from Cohen's effect size calculation showed that the treatment condition created a large effect size of 1.04 and 1.11 on word-form acquisition and retention, respectively. However, when this vocabulary gain is evaluated, from a pedagogic standpoint, in terms of the actual number of target words that the learners recalled the form of, it proves to be relatively small.

Friedman tests were carried out for each learning condition to examine the difference in the learners' vocabulary test scores across the three testing times. While these scores remained the same across the three testing times in the case of the control condition, a significant score improvement was found for the treatment condition: $X^2(2, N = 18) = 6.86$ ($p < .05$). In this condition, the learners' scores in the immediate and delayed post-tests were nearly identical and markedly higher than those in the pre-test. Again, however, this difference only reached significance when the Bonferroni post-hoc correction was not applied: $Z = 2.67$ ($.017 < p < .05$).

To compare the size of word-meaning acquisition and retention between the groups, two Mann Whitney tests were run. The learners in the treatment condition were indeed found to make significantly larger word-meaning acquisition and retention than those in the control condition with $U = 254$ ($p < .01$) and $U = 261$ ($p < .01$), respectively.

Association between Vocabulary Gain and Text Comprehension

The data on the text comprehension measure were only provided by 28 out of the 32 learners in this study as the test validation (see the Methodology Chapter) showed that the

text comprehension levels by four learners were not effectively gauged by the current test. For the remaining learners, they were found to obtain a mean score of 9.03 (SD 1.83) out of the maximum score of 15 points (60.20%). Their test scores ranged from 7 (46.67%) to 13 (86.67%) points.

To examine the direction, strength and significance of the association between vocabulary gain and text comprehension, Pearson correlation coefficients were calculated. On the word-form recall test, the coefficients between the text comprehension scores on the one hand and the acquisition and retention scores on the other hand were the same: $r = .58$ ($p < .05$). These coefficients in the case of the word-meaning measure were $.58$ ($p < .05$) and $.51$ ($p < .05$), respectively. Thus, the more successfully the learners understood and remembered the video content, the more vocabulary knowledge they tended to pick up.

Associations between Vocabulary Gain and other Factors of Interest

Pearson correlation coefficients were also computed to gauge the direction, strength, and significance of the associations between the vocabulary gain in the treatment condition on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input texts, and the usefulness of pictorial and verbal clues in these texts for word-meaning inference on the other hand. The results are given in Table 21 below:

Table 21: Association between vocabulary gain in NV and factors of interest

		L2 proficiency	Vocabulary knowledge	Word occurrences	Verbal clues	Pictorial clues
Form	Acquisition	.65*	.21	.45*	.39	.34
recall	Retention	.65*	.21	.45*	.39	.34
Meaning	Acquisition	.51*	.44	.36	.44*	.12
recall	Retention	.51*	.46	.33	.47*	.14

* $p < .05$

As can be seen from this table, the size of vocabulary gain in this learning condition was positively correlated with the learners' prior L2 proficiency. In addition, the word-form gain tended to be dependent on the frequency of word occurrences in the videos. Meanwhile the word-meaning gain was positively associated with the usefulness of verbal clues in these videos for word-meaning inference. The relationships between the size of vocabulary gain and the other factors of interest were all positive, but consistently far from linear.

Summary

The report above shows that L2 learners were able to make a statistically significant vocabulary gain from the longest video and all three videos in this treatment condition. On average, they picked up half a target word from the longest video and one target word from all three videos. The vocabulary gain from all three videos was positively correlated with the learners' text comprehension. A positive correlation was also found for the relationship between this vocabulary gain and their prior L2 proficiency. If the size of word-form gain was associated with the frequency of word occurrences in the input materials, that of word-meaning gain tended to depend on the usefulness of verbal clues in the videos for word-meaning inference.

Study 5: Adding an Output Task into the Narrow Viewing Procedure (NVS)

This study measured the size of vocabulary gain from a learning condition where L2 learners were first required to view either the first or the second half of the three-video set in Study 4, report the video content to a classmate who had not viewed the same half yet in an interactional summary activity, and then move on with the other half (i.e., the NVS condition). The association between this vocabulary gain and the learners' text comprehension were also examined in the present study. So were the relationships between this vocabulary gain on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input materials, and the usefulness of verbal and pictorial clues in the input materials for word-meaning inference on the other hand. The research questions investigated in this study were as below:

- (m) To what extent do L2 learners incidentally pick up the form and meaning of novel words from viewing either the first or the second half of the three-video set in Study 4, reporting the video content to a classmate who had not viewed the same half yet in an interactional summary activity, and then moving on with the other half?*
- (n) Is this vocabulary gain associated with the learners' text comprehension?*
- (o) Is this vocabulary gain associated with the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the videos, and the usefulness of verbal and pictorial clues in these videos for word-meaning inference?*

Word-form Gain

Table 22 presents the descriptive statistics for the scores that the learners in the control group and the treatment group achieved on the word-form recall measure. Again, after removing the target words that were already known to more than 10% of the learners, the data on this measure were now related to only 25 out of the 40 target words.

Table 22: Word-form gain in NVS – 25 target words (max. = 25 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	14	0.36	0.50	14	0.50	0.52
Post-test	14	0.50	0.65	14	2.29	1.49
Delayed post-test	14	0.43	0.51	14	2.21	1.55
Acquisition	14	0.14	0.36	14	1.79	1.19
Retention	14	0.07	0.26	14	1.71	1.25

To ensure that there was no significant difference in the learners' prior knowledge of the form of the target words between the groups, a Mann Whitney test was carried out with their pre-test scores. The result indeed showed that the learners began this experiment with a similar amount of knowledge regarding the form of the target words: $U = 105$ ($p = .76$).

As the descriptive statistics in Table 22 show, the learners in the treatment condition picked up about two target words on this measure. Cohen's d values were computed to gauge the effect size of this condition on word-form gain. This condition was found to bring about a large effect size of 1.88 on word-form acquisition and of 1.82 on word-form retention.

Friedman tests were implemented to examine whether the learners' scores in each learning condition significantly improved from the pre-test to the post-tests. In the control condition, these scores remained relatively similar across the three testing times: $X^2(2, N = 14) = 0.32$ ($p = .85$). Meanwhile, a significant score growth was observed in the case of the treatment condition: $X^2(2, N = 14) = 16.82$ ($p < .001$). The results from the Wilcoxon Signed-Rank tests again showed that the two post-test scores were not significantly different ($Z = 0.23$, $p = .82$), but both significantly surpassed the pre-test scores even when the Bonferroni post-hoc correction was used with the same $Z = 3.60$ ($p < .017$). Mann Whitney tests were used to compare the size of word-form gains between the conditions. The treatment condition was found to bring about significantly larger word-form acquisition and retention than the control condition with the same U value of 196 ($p < .0001$).

Word-meaning Gain

The descriptive statistics for the scores that the learners in the two groups obtained on the word-meaning recall measure are summed up in Table 23. The data on this measure were now related to only 29 out of the 40 target words.

Table 23: Word-meaning gain in NVS – 29 target words (max. = 29 points)

Scores	Control condition			Treatment condition		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Pre-test	18	0.50	0.71	18	0.44	0.62
Post-test	18	0.50	0.71	18	2.00	1.03
Delayed post-test	18	0.50	0.71	18	2.00	1.03
Acquisition	18	0.00	0.00	18	1.56	0.78
Retention	18	0.00	0.00	18	1.56	0.78

A Mann Whitney test was run to compare the learners' pre-test scores between the groups. No difference was found: $U = 153$ ($p = .79$). Thus, the learners in two groups had an equal starting point regarding their knowledge of the meaning of the target words before this experiment started.

On average, the learners in the treatment condition successfully recalled the meaning of two target words in the post-tests. To measure the effect size of this condition on word-meaning gain, Cohen's d values were calculated. The results showed that this condition created a large effect size of 2.83 on both word-meaning acquisition and retention.

To examine the difference in the learners' vocabulary test scores across the three testing times, Friedman tests were conducted for each learning condition. No difference was found in the control condition: $X^2(2, N = 18) = 0.00$ ($p = 1.00$). Meanwhile, a significant difference was found in the case of the treatment condition: $X^2(2, N = 18) = 21.36$ ($p < .0001$). The results from the Wilcoxon Signed-Rank tests again indicated that the two post-test scores were exactly the same and significantly surpassed the pre-test scores even when the Bonferroni post-hoc correction was used: $Z = 3.52$ ($p < .017$).

Mann Whitney tests were carried out to compare the size of word-meaning gains between the conditions. The results showed that the treatment condition resulted in significantly better word-meaning acquisition and retention than the control condition with the same U value of 295 ($p < .0001$).

Association between Vocabulary Gain and Text Comprehension

The data on the text comprehension measure were only provided by 30 out of the 32 learners in this study as the test validation (see the Methodology Chapter) indicated that the text comprehension levels of two learners were not successfully gauged by the current test and therefore their test scores should be removed from the final analyses. For the remaining learners, they were found to obtain a mean score of 11.07 (*SD* 1.20) out of the maximum score of 15 points (78%). Their test scores ranged between 9 (60%) to 13 (86.67%) points.

To gauge the direction, strength, and significance of the association between vocabulary gain and text comprehension, Pearson correlation coefficients were computed. The coefficients between the text comprehension scores on the one hand and the word-form acquisition and retention scores on the other hand were .54 ($p < .05$) and .57 ($p < .05$), respectively. These coefficients in the case of the word-meaning recall measure were .61 ($p < .01$) and .71 ($p < .001$). Therefore, the better the text comprehension the learners obtained from their viewing procedure, the more the vocabulary knowledge they tended to gain.

Associations between Vocabulary Gain and other Factors of Interest

Table 24 sums up the results from Pearson correlation calculations for the relationships between the vocabulary gain in the treatment condition on the one hand and the learners' prior L2 proficiency and vocabulary knowledge, the frequency of word occurrences in the input materials, and the usefulness of verbal and pictorial clues in these materials for word-meaning inference on the other hand:

Table 24: Association between vocabulary gain in NVS and factors of interest

		L2	Vocabulary	Word	Verbal	Pictorial
		proficiency	knowledge	occurrences	clues	clues
Form	Acquisition	.38	.31	.58**	.26	.23
recall	Retention	.34	.33	.62**	.30	.28
Meaning	Acquisition	.51*	.23	.48**	.39*	.13
recall	Retention	.50*	.21	.47*	.41*	.12

* $p < .05$; ** $p < .01$

As can be seen from this table, the more often a target word occurred in the videos, the more likely it was picked up. In addition, the size of word-meaning gain was also found

to be positively correlated with the learners' prior L2 proficiency and the usefulness of verbal clues in the videos for word-meaning interpretation.

Summary

The treatment condition in this study was also found to bring about a statistically significant vocabulary gain. The learners, on average, learnt two target words from this viewing procedure. This vocabulary gain was positively associated with the learners' text comprehension and the frequency of word occurrences in the videos. In addition, the size of word-meaning gain was also positively correlated with the learners' prior L2 proficiency and the usefulness of verbal clues in the videos for word-meaning inference.

Study 6: A Comparison of Effect Sizes on Vocabulary Gain

The aim of this study was to compare the effect sizes of all treatment conditions on vocabulary gain and find out the factor or factors that could explain the difference in these effect sizes (if any). Put differently, this study addressed the two research questions below:

(p) Is there any difference in the effect sizes of all treatment conditions on vocabulary gain?

(q) If yes, which factor(s) in the conditions account(s) for this difference?

A Comparison of Effect Sizes on Vocabulary Gain

Cohen's d values of all treatment conditions as found from Study 1 to Study 5 were integrated into the meta-analysis program in the Exploratory Software of Confidence Intervals (ESCI) (Cumming, 2012) to rank these conditions according to their true effect sizes on vocabulary gain. It should be noted that Study 4 included two different datasets – one for the vocabulary gain made on the 20 target words from the longest video and the other for the lexical uptake obtained out of the 40 target words from all three videos. In the following comparisons, the vocabulary gains from these two datasets will be referred to as NV1 and NV2, respectively.

These effect sizes on word-form acquisition and retention are plotted in Graph 1 and 2 below:

Graph 1: Effect sizes on word-form acquisition **Graph 2:** Effect sizes on word-form retention

The graphs above show that all treatment conditions in this research project had a large effect size on word-form gain. These effect sizes are ascendingly ranked as NV2 < OV < NV1 = RV < RVS < NVS in the case of the word-form acquisition and NV1 = NV2 < OV < RV < RVS < NVS in the case of the word-form retention.

The effect sizes of these conditions on the word-meaning gain are also large, as can be seen from Graph 3 and 4. Their ascending ranking orders are OV < NV1 < NV2 < RV < RVS < NVS and OV < NV1 = NV2 < RV < RVS < NVS for the word-meaning acquisition and retention, respectively.

Graph 3: Effect sizes on word-meaning acquisition **Graph 4:** Effect sizes on word-meaning retention

It should be noted that the learners in the OV condition, the RV condition, and the RVS condition viewed the same video and were exposed to the same target words. Therefore, it was worth comparing the effect sizes on vocabulary gain among these conditions. According to the comparisons above, the ascending ranking order of these conditions is $OV < RV < RVS$. Similarly, the learners in the NV condition (on the 40-target-word or NV2 dataset) and the NVS condition also watched the same videos and were exposed to the same target words. The ranking order of their effect sizes is NV (or, to be more precise, NV2) < NVS. The learners in the NV condition (on the 20-target-word or NV1 dataset) were also expected to obtain a larger vocabulary gain than those in the OV condition as the former might have increased their topic familiarity and lexical coverage of the longest video thanks to viewing beforehand the two shorter videos which share almost the same topic as that video. In most cases, the comparisons above support this expectation: NV (or, to be more precise, NV1) > OV.

Explanatory Factors for the Difference in these Effect Sizes

Vocabulary gains in the OV, the RV, and the RVS condition

Association between vocabulary gain and text comprehension

Another ESCI analysis was carried out to compare the degrees of text comprehension across the three learning conditions: OV, RV and RVS. The result is presented in Funnel Plot 3.1:

Funnel Plot 1: Text comprehension in OV, RV and RVS

According to this funnel plot, the ascending ranking order of the text comprehension levels obtained in these conditions is $OV < RV < RVS$.

To examine the association between vocabulary gain and text comprehension in these conditions, the data from the three groups were first combined. Subsequently, Pearson correlation coefficients were computed. The results showed that the text comprehension

scores were positively correlated with the word-form acquisition and retention scores with the same coefficient of .71 ($p < .0001$). These coefficients in the case of the word-meaning acquisition and retention scores were .69 ($p < .0001$) and .71 ($p < .0001$), respectively. Thus, the more successfully the learners understood the video content, the larger the vocabulary gain they tended to make.

To investigate whether it was (a) the difference in text comprehension or (b) the other difference in learning conditions that resulted in the difference in vocabulary gains among these conditions, the two variables (a) and (b) were integrated into a multiple regression model to gauge their predictive power for the size of vocabulary gains.

On the word-form recall measure, this model was found to explain 50.69% of the variance in the acquisition scores: $F(2, 37) = 22.07$ ($p < .0001$). Text comprehension significantly predicted the size of word-form acquisition: $Beta = .29$, $t = 3.56$ ($p < .001$). The predictive power of the other difference in learning conditions, however, fell short of the required significance level: $Beta = .36$, $t = 1.51$ ($p = .14$). This model also accounted for 53.25% of the variance in the retention scores: $F(2, 37) = 24.35$ ($p < .0001$). Both text comprehension and the other difference in learning conditions significantly predicted the size of word-form retention with $Beta = .26$, $t = 3.32$ ($p < .01$) and $Beta = .49$, $t = 2.06$ ($p < .05$), respectively.

On the word-meaning measure, this model explained 50.23% of the variance in the acquisition scores: $F(2, 51) = 27.74$ ($p < .0001$). Text comprehension significantly predicted the size of word-meaning acquisition ($Beta = .22$, $t = 2.10$, $p < .0001$). So did the other difference in learning conditions ($Beta = .32$, $t = 4.52$, $p < .05$). This model also accounted for 53.58% of the variance in the retention scores: $F(2, 51) = 31.58$ ($p < .0001$). Again, both text comprehension and the other difference in learning conditions significantly predicted the size of word-meaning retention with $Beta = .21$, $t = 2.52$ ($p < .0001$) and $Beta = .36$, $t = 4.58$ ($p < .01$), respectively.

As expected, the difference in the text comprehension levels obtained by the learners in these conditions accounted for much of the difference in their vocabulary gains. However, the results from the Multiple Regression analyses above also indicated that there were other explanatory factors for this difference.

Association between vocabulary gain and word repetition

As reported in Study 1, 2 and 3, one possible factor that may also explain the difference in vocabulary gains across the OV condition, the RV condition, and the RVS condition is the difference in the frequency at which the learners encountered the target words in the input material. As the learners in the RV condition and the RVS condition viewed the video twice, they encountered the target words twice as often as those in the OV condition who viewed this video only once.

The data from these conditions were first combined and Pearson correlation coefficients were subsequently computed for the relationship between the frequency of word encounters in the input material and the size of vocabulary gains. Word repetition was indeed positively correlated with the word-form acquisition and retention with a coefficient of .47 ($p < .01$) and .50 ($p < .001$), respectively. These coefficients in the case of the word-meaning acquisition and retention were .45 ($p < .001$) and .44 ($p < .001$). Therefore, the more often a target word was encountered in the input material, the more likely it was picked up.

To investigate whether it was (a) the difference in the frequency of word encounters or (b) the other difference in learning conditions that explained the difference in vocabulary gains across these conditions, the two variables (a) and (b) were integrated in another multiple regression model to gauge their predictive power for the size of vocabulary gains. On the word-meaning recall measure, this model was found to explain 36.14% and 42.21% of the variance in the acquisition scores ($F(2, 37) = 12.60, p < .0001$) and the retention scores ($F(2, 37) = 15.97, p < .0001$), respectively. The other difference in the learning conditions significantly predicted the size of word-form acquisition ($Beta = 1.29, t = 3.29, p < .01$) and word-form retention ($Beta = 1.43, t = 3.81, p < .001$). However, the predictive power of the frequency of word encounters for the size of word-form acquisition and retention was non-significant with $Beta = -.64, t = -.95 (p = .35)$ and $Beta = -.79, t = -1.21 (p = .23)$, respectively.

On the word-meaning measure, this model also accounted for 31.41% and 37.45% of the variance in the acquisition scores ($F(2, 51) = 13.13, p < .0001$) and the retention scores ($F(2, 51) = 16.86, p < .0001$), respectively. Again, the other difference in learning conditions significantly predicted the size of word-meaning acquisition and retention, with $Beta = .94, t = 3.30 (p < .01)$ and $Beta = .11, t = 4.14 (p < .001)$, respectively. Meanwhile, the difference

in the frequency of word encounters could not predict either word-meaning acquisition ($Beta = -.44, t = -.90, p = .37$) or word-meaning retention ($Beta = -.72, t = -1.56, p = .13$).

Taken together, the frequency of word encounters in the input material was positively associated with the size of vocabulary gains in these conditions. However, the effect of treatment conditions was such that it overrode the explanatory power of this association.

Association between vocabulary gain and use of target words

Regarding the difference in vocabulary gains between the RV condition and the RVS condition, there was another factor that may also play a role. That is the way the learners in the RVS condition dealt with the target words in their summary activity. It should be also noted that no target word use was detected from the learners' note sheets in the RV group.

To examine whether the learners in the RVS condition might have enhanced their vocabulary gain thanks to their use of the target words in this output task, their summaries were transcribed and screened for the presence of target words. Altogether I found four out of the 18 target words in my collection of the summaries (*lure* (x9 tokens), *bait* (x1), *bioluminescence* (x2), and *saber* (x1)). Eleven out of the 32 learners in this group used one or more target words in their summaries. The total number of times these words occurred in these summaries was 14 (one learner used one target word three times and another target word once while the other ten learners used only one target word and did so only once). These words were all unknown prior to the treatment, according to the pre-test data, and yet they were used in the summaries in grammatically and semantically appropriate ways.

It is also interesting to note that the syntagmatic contexts in which these target words were used all differed from the original phrasings in the video. Below is one example:

"This is an optical lure that I have used... And I cannot claim for sure that the optical source brought it in, because there is bait right there." (Original context)

"In two examples, she used an optical lure. When this optical lure started producing light, four or five fish came along and they started to produce light and together lighting up for communication. And then a shark came. She said she didn't know it was because of this light or the bait that she put it in the optical lure that lured the shark." (RVS320)

In these summaries, I also found 36 instances of conspicuously long (more than 2.00 seconds) pauses where a target word would have been appropriate. All these pauses occurred in the middle of a clause, which is a likely sign that the learners were looking for the target

words (Tavakoli, 2011), but then failed to retrieve them from memory. After the pauses, the learners would either resort to a substitution, like “*bio-illumination*” for “*bioluminescence*”, or to a paraphrase, as in “*They set [a long pause] a machine; the machine can produce light and therefore it attracts many fish like an octopus and shark*”, presumably when searching for “*optical lure*”. Interestingly, 13 of these 36 episodes suggesting retrieval failure were followed by successful word recall in the post-tests. It is conceivable that the retrieval failures prompted these learners to take notice of these elusive target words as they listened to the text a second time.

Thus, there was evidence that the learners in the RVS condition used the target words in their summaries and by doing so created a durable memory of these lexical items. This summary activity also appeared to have prompted the learners to notice gaps in their lexical resources which they then filled in their second viewing. The combined effect of this use and noticing may have enlarged the size of their lexical uptake.

Vocabulary gains in the NV and the NVS condition

Association between vocabulary gain and text comprehension

Another ESCI analysis was also run to compare the degrees of text comprehension between the NV condition and the NVS condition. The result is given in Funnel Plot 2:

Funnel Plot 2: Text comprehension in NV and NVS

This funnel plot shows that the NVS condition resulted in a much better text comprehension level than the NV condition. Again, Pearson correlation coefficients were computed to gauge the direction, strength, and significance of the association between vocabulary gain and text comprehension in these conditions. The coefficients between the text comprehension scores on the one hand and the word-form acquisition and retention scores on the other hand were .65 ($p < .001$) and .66 ($p < .001$), respectively. These coefficients in the case of the word-meaning recall measure were .58 ($p < .001$) and .54 ($p < .001$), respectively.

.01). Again, the better the text comprehension the learners obtained from their viewing procedure, the larger the vocabulary gain they were likely to achieve.

To examine whether (a) the difference in text comprehension or (b) the other difference in learning conditions resulted in the difference in vocabulary gains between the conditions, the variables (a) and (b) were integrated in a multiple regression model to gauge their predictive power for the gain size. On the word-form recall test, this model explained 36.98% of the variance in the acquisition scores ($F(2, 23) = 8.33, p < .01$) and 39.34% of the variance in the retention scores ($F(2, 23) = 9.11, p < .01$). Text comprehension significantly predicted the size of word-form acquisition and retention with $Beta = .36, t = 2.98 (p < .01)$ and $Beta = .40, t = 3.26 (p < .01)$, respectively. Meanwhile, the other difference in learning conditions failed to predict the size of word-form acquisition ($Beta = -.01, t = -.02, p = .99$) and retention ($Beta = -.11, t = -.24, p = .82$). On the word-meaning recall test, this model was found to account for 37.35% of the variance in the acquisition scores: $F(2, 29) = 10.24 (p < .001)$. Text comprehension significantly predicted the size of word-meaning acquisition ($Beta = .30, t = 3.16, p < .01$). Meanwhile, the predictive power of the other difference in learning conditions narrowly missed the required significance level: $Beta = .56, t = 1.95 (p = .06)$. This model also explained 41.93% of the variance in the retention scores: $F(2, 29) = 12.19 (p < .001)$. Both text comprehension and the other difference in learning conditions significantly predicted the size of word-meaning retention with $Beta = .23, t = 2.76 (p < .01)$ and $Beta = .75, t = 2.89 (p < .01)$, respectively. Together, the difference in text comprehension in these learning conditions could explain the difference in their vocabulary gains.

It should be also noted that the learners in the NV condition were required to watch the three videos with a view to answering the content questions. Meanwhile, those in the NVS condition watched either the two shorter videos or the longest video, summarized the content to a classmate who had not watched the same video(s) in an information-gap activity, and then moved on with the other video(s). A pertinent question here was whether *giving* a summary of the video(s) after viewing or *receiving* a summary of the video(s) before viewing led to better text comprehension as compared to merely *viewing* the videos with anticipation of the content questions. To answer this question, the text comprehension scores were first split into two subsets: those for the two shorter videos and those for the longest video. In each of these subsets, the scores were then separated for the learners who merely watched the videos ($n = 28$), those who watched the video(s) and then gave a summary of the video content ($n = 15$), and those who listened to a summary of the video content before they

actually watched the video(s) themselves ($n = 15$). Hereafter, the learners in the three groups will be referred to respectively as *viewers only*, *summary givers* and *summary receivers*.

Mann Whitney tests were run to compare the text comprehension scores across the groups. For the two shorter videos, both summary givers and receivers obtained significantly higher text comprehension scores than viewers only, with $U = 172$ ($p < .01$) and $U = 163$ ($p < .001$), respectively. These scores, however, did not significantly differ between summary givers and receivers: $U = 59$ ($p = .52$). For the longest talk, again both summary givers and receivers significantly outperformed viewers only on this measure, with $U = 157$ ($p < .05$) and $U = 152$ ($p < .05$), respectively. The performance of summary givers and receivers on this measure was not significantly different: $U = 55$ ($p = .73$). Thus, having learners give a summary of the video content after viewing and receive a summary of the video content before viewing were both found to bring about significantly better text comprehension than having them merely view the videos.

The finding above logically leads to another question, that is, whether giving a summary of the video content after viewing and receiving a summary of the video content before viewing was also conducive to incidental vocabulary acquisition. To answer this question, all the summaries produced by the learners in the NVS condition were first transcribed ($N = 30$). Two experienced EFL/ESL teachers were invited to watch the videos, read their transcripts, and then holistically rate the quality of the learners' summaries on a ten point scale. Their independent ratings were strongly correlated, with $r = .81$ ($p < .0001$). The means of their ratings were used to calculate two Pearson correlation coefficients with the vocabulary gain obtained by those who gave and those who received the summary, respectively. In the former, the vocabulary gain was computed only on the target words that occur in the video that the learners watched before making the summary. In the latter, it referred to the gain made on the target words which the learners encountered in the video after they received the summary. The outcome is presented in Table 25:

Table 25: Association between vocabulary gain and summary quality

Vocabulary measure	For	Correlation coefficients
Form recall ($n = 14$)	Summary givers	Acquisition: $r = -.24$
		Retention: $r = -.22$
	Summary receivers	Acquisition: $r = .51$
		Retention: $r = .62^*$

Meaning recall ($n = 16$)	Summary givers	Acquisition: $r = -.37$
		Retention: $r = -.26$
	Summary receivers	Acquisition: $r = .64^{**}$
		Retention: $r = .54^*$

* $p < .05$; ** $p < .01$

Although the required significance level was not reached, there was a clear tendency for a negative correlation between the learners' summary quality and their own vocabulary gain. In other words, the better their summary, the smaller their lexical uptake. By contrast, their summary quality was positively, moderately, and significantly correlated with their classmates' vocabulary gain made from watching the summarized video(s). It should be noted that all the learners in the NVS condition benefited from the classroom procedure, because they were all summary receivers at some point of the lesson.

Therefore, although giving a summary of the video content after viewing and receiving a summary of the video content before viewing both led to significantly better text comprehension than merely viewing the videos, it was receiving a summary of the video content before viewing that helped the learners in the NVS condition to enlarge their vocabulary gain.

Association between vocabulary gain and use of target words in the summary activity

From all the summaries produced by the learners in the NVS condition ($N = 30$), I found only two out of the 29 target words (6.90%) that were used: “*lure*” and “*plunge*”. Five learners used the word “*lure*” for a total number of six times while only one used the word “*plunge*” twice. In all instances, such use was grammatically correct and semantically appropriate as seen below:

“In the video, she first showed a video about the glowing of animals under the ocean when she first dived, plunged, she first took a plunge and then she was amazed by it. She made a point that those glowing animals used these lights intentionally, just like a scream for help, to grasp attention or to lure a prey or predator or something like that.” (NVS515)

It was interesting to note that this learner NVS515 even attempted to experiment with using these words as verbs – a grammatical function that did not occur in the original text. However, when he did not feel confident enough with this use, he brought the use of the word “*plunged*” as a verb back to the original grammatical function as a noun in the video “*took a plunge*”.

A cross-check with the pre-test and post-test scores of the learners who actually used these words in their summaries indicated that none of them knew the meaning of these words before watching the videos (as evidenced in their pre-test performance), but they did use these words in their summaries and remembered them in both post-tests.

Regarding the benefit of interaction for incidental vocabulary acquisition, from the 30 summaries, I found only one episode where two learners negotiated the meaning of the word *lure*:

NVS519: “...In the last part of the TED Talk, the speaker said about a device – the optical lure – it’s a kind of camera.”

NVS520: “Did you say optical...”?

NVS519: “Lure. I don’t know what exactly it means. I have seen an illustrated image in the screen and it’s a kind of camera, a deep sea webcam. It can be put at 3,000 feet and it can record creatures’ images.”

Although NVS519 tried to explain one shade of the meaning of the target word *lure* as “it’s a kind of camera”, this shade of meaning does not fully denote what the speaker intended to convey in the video. NVS519 also admitted to her interlocutor NVS520 that she had not got the precise meaning of this word yet. Their test scores suggested that both of them were still unable to figure out the contextual meaning of this word even after watching this video. In other words, their negotiation of the word meaning above did not lead to uptake of the precise meaning of this item.

All the analyses above indicate that the learners in the NVS condition did use and/or negotiate the meaning of some newly met words in their output task. However, the number of such instances was too limited to make a significant contribution to their vocabulary gain. These findings are consistent with the results from the Multiple Regression analyses above.

Vocabulary gains in the OV and the NV condition

On the dataset of the 20 target words from the longest video, the learners in the NV condition outperformed those in the OV condition in the word-form immediate post-test, the word-meaning immediate post-test, and the word-meaning delayed post-test. Meanwhile, the learners in the OV condition only scored higher in the case of the word-form delayed post-test. The generally better vocabulary gain in the NV condition may be attributed to the effect of viewing the two shorter videos before the longest video. As the three videos have similar topics, this viewing procedure was deemed to help the learners to increase their topic familiarity and lexical coverage when they came to the longest video. This, in turn, was expected to foster their comprehension of the longest video.

As the learners in these two conditions had two different text comprehension tests, I could not make a direct comparison of the text comprehension scores between the groups. However, an ESCI analysis can provide a relative comparison of the effect sizes that these conditions produced on text comprehension. The result of this analysis is presented in Funnel Plot 3:

Funnel Plot 3: Text comprehension in the OV and the NV condition

According to this plot, the learners in the NV condition indeed obtained better text comprehension than those in the OV condition. There were five overlapping questions between the two text comprehension tests that the OV and the NV groups took. Therefore, it was worth comparing the scores on these five items between the groups. The descriptive statistics for these scores are summed up in Table 26:

Table 26: Text comprehension in the OV and the NV condition (max. = 5)

OV ($N = 31$)	NV ($N = 28$)
1.23 ($SD\ 0.99$)	2.25 ($SD\ 1.11$)

A Mann Whitney test was carried out to gauge the difference in the text comprehension scores between the groups. The result shows that the learners in the NV condition scored significantly higher than those in the OV condition: $U = 641$ ($p < .01$). To examine the association between the text comprehension scores and the vocabulary gain scores between the two conditions, their data were first combined and Pearson correlation coefficients were then computed to get the direction, strength and significance of this association. The text comprehension scores were found to be positively correlated with the word-form acquisition scores ($r = .63$, $p < .001$) and the word-form retention scores ($r = .77$, $p < .0001$). These coefficients in the case of the word-meaning acquisition and retention scores were $r = .64$ ($p < .0001$) and $r = .73$ ($p < .0001$), respectively.

To verify whether it was (a) this better text comprehension or (b) the other difference in learning conditions that gave the NV group an advantage for incidental vocabulary uptake, these two variables were again incorporated into a multiple regression model to gauge their predictive power for the size of vocabulary gains. On the word-form recall measure, this

model explained 36.44% and 55.51% of the variance in the acquisition scores ($F(2, 22) = 7.88, p < .01$) and the retention scores ($F(2, 22) = 15.97, p < .0001$), respectively. The text comprehension scores significantly predicted the size of word-form acquisition ($Beta = .25; t = 2.91, p < .01$) and retention ($Beta = .31; t = 4.60, p < .001$). Meanwhile, the contribution of the other difference in learning conditions to the explanatory power of this model was non-significant with $Beta = .17, t = 0.89 (p = .39)$ for the word-form acquisition and $Beta = .08, t = 0.56 (p = .59)$ in the case of the word-form retention. On the word-meaning measure, this model also accounted for 37.09% of the variance in the acquisition scores ($F(2, 31) = 10.73, p < .001$) and 49.55% of the variance in the retention scores ($F(2, 31) = 17.20, p < .0001$). Again, the text comprehension scores were found to contribute the most to the predictive power of this model with $Beta = .36, t = 4.34 (p < .001)$ for the word-meaning acquisition and $Beta = .39, t = 5.30 (p < .0001)$ in the case of the word-meaning retention. The contribution of the other difference in learning conditions was negligible with $Beta = -.09; t = -0.47 (p = .64)$ and $Beta = -.02; t = -0.10 (p = .93)$ for the word-meaning acquisition and retention, respectively.

It should be also noted that most of the target words picked up by the learners in the NV condition are from the longest video rather than from the two shorter videos. Put differently, the learners in this condition actually obtained more vocabulary knowledge from the longest video than from the two shorter videos. This is clearly evidenced by a very small discrepancy between their mean scores out of the 20 target words from the longest video and out of the 40 target words from all three videos, as can be seen in Table 27 below:

Table 27: Vocabulary gain in the NV condition

20 target words	40 target words
Word-form acquisition: 0.57	Word-form acquisition: 0.79
Word-form retention: 0.50	Word-form retention: 0.79
Word-meaning acquisition: 0.55	Word-meaning acquisition: 0.67
Word-meaning retention: 0.55	Word-meaning retention: 0.56

This also suggests that viewing the two shorter videos before the longest one might have increased the learners' topic familiarity and lexical coverage when they came to the longest video and therefore could have fostered their understanding of and incidental vocabulary gain from this video. However, the fact that these learners picked up more words from the longest video than from the two shorter ones may be also due to the longest video being sequenced the last in their viewing procedure and thus relatively close to the time of the vocabulary post-test. Accordingly, the learners may have remembered words in this video

better than lexical items in the two shorter videos, which were arranged at the beginning of their viewing procedure.

To recap, the NV condition brought about a larger vocabulary gain out of the 20 target words from the longest video than the OV condition. This better vocabulary gain could be attributed to the positive effect of first viewing the two shorter videos, which share almost the same topic as the longest video. This narrow viewing procedure was found to noticeably reduce the number of potentially unfamiliar words that the learners encountered in the longest video (please refer to the Methodology Chapter for a full analysis of the lexical overlap) and also enhance their comprehension of the video content (as analysed above). As a result, this procedure may have created a favourable condition for incidental vocabulary uptake from this video.

Summary

The learners in the RV condition obtained a larger vocabulary gain than those in the OV condition thanks to their second viewing, where they could considerably enhance their text comprehension. The RVS condition also led to a more pronounced vocabulary gain than the RV condition. This better lexical uptake was due to the summary activity, where the learners could (a) boost their text comprehension, (b) use the target words and therefore foster their memorization of these lexical items, and (c) notice gaps in their lexical repertoire and then fill these gaps in their second viewing. Regarding the difference in the effect sizes of the NV condition and the NVS condition on vocabulary gain, text comprehension was again found to be the main explanatory factor. To be more specific, receiving a summary of the video content before viewing facilitated L2 listening comprehension in the NVS condition and therefore also fostered incidental vocabulary acquisition. The learners in the NV condition were also found to make a larger vocabulary gain from the longest video than those in the OV condition thanks to the benefit of narrow viewing or, to be more precise, viewing the two other videos of virtually the same topic before the longest one. This procedure helped to improve their comprehension of this video and therefore may have benefited lexical uptake from this video.

Study 7: A Qualitative View into the Benefits of the Output Tasks for Vocabulary Gain

In Study 3 and 5, the monologue summary activity in the RVS condition and the interactional summary activity in the NVS condition were both found to benefit incidental vocabulary uptake. However, due to the quantitative nature of these studies, how these summary activities actually fostered lexical gain was not yet understood. This limitation was

removed in this study, where I used a follow-up interview with a different group of eight learners (four had undertaken the RVS condition and the other four had followed the NVS condition) about the cognitive and metacognitive processes that they had used to deal with the summary activities and how these processes had affected their text comprehension and incidental vocabulary uptake. This study addressed the following questions:

- (r) *To what extent do L2 learners incidentally pick up the form and meaning of novel words from the RVS condition and NVS condition?*
- (s) *What cognitive and metacognitive processes do they use to deal with the summary activities in these conditions and how do these processes affect their text comprehension and incidental vocabulary acquisition?*

Benefits of the Monologue Summary Activity in the RVS Condition

Quantitative results

Table 28 presents the learners' raw scores on the text comprehension measure and the difference between their pre-test and immediate post-test scores on the vocabulary measures. It should be noted that the learners in this study did not take the delayed vocabulary post-test:

Table 28: Text comprehension and Vocabulary gain in the RVS Condition

Learners	Text comprehension	Vocabulary Gain	
		Form Recall	Meaning Recall
RVS601	10		3/18
RVS602	12		5/14
RVS603	12	4/17	
RVS604	10	5/15	

On the text comprehension measure, two learners (RVS601, RVS604) obtained a score of 10 out of the maximum score of 15 points (67%) while the other two (RVS602, RVS603) achieved a score of 12 (80%). After removing the target words that the learners already knew prior to the experiment, RVS601 and RVS602 were found to make a gain of 3 (17%: *bioluminescence*, *breath-taking*, and *lure*) and 5 (36%: *bait*, *enchanted*, *foolhardy*, *lure*, and *unobtrusive*) out of 18 and 14 target words on the word-meaning recall test, respectively. Meanwhile, RVS603 and RVS604 obtained a gain of 4 (24%: *bioluminescence*, *breath-taking*, *lure*, and *mesh*) and 5 (33%: *contemplative*, *junky*, *lure*, *plunge*, and *unobtrusive*) out of 17 and 15 target words on the word-form recall test, respectively.

Together, the learners in this learning condition attained a noticeable degree of both text comprehension and vocabulary gain from their assigned viewing procedure.

Qualitative results

All learners in this condition reported that the summary activity prompted them to reprocess perceived information in their notes, which, in turn, improved their understanding of the listening content. RVS602, for example, described this benefit in an analogy: *“This summary activity is just like a way for me to step back to re-evaluate what I have heard and to see the whole picture of the given message. In this way, I can improve my understanding of the video content considerably”*. In this information reprocessing, the learners carried out different cognitive activities. For example, they attempted to distinguish the main ideas of the given video from the supporting ideas and factual information: *“In the video, the presenter starts her talk by inviting her audience to an alien world that is our ocean. However, I think this information is too specific so I did not use it in my summary. I tried to fasten my focus on main ideas only”* (RVS601) or *“This summary activity required me to skim my notes again for the main ideas. Such detailed information as a coral bush is 3,000 years old will not be used in my summary”* (RVS602). They also tried to organize these main ideas in a coherent summary: *“My notes look rather messy as I listed information in different parts of the note sheet. As I am required to summarize the video content, I have to group ideas in a particular organization”* (RVS601) or *“This summary activity helped me to look at my notes again, and then to organize ideas in an easy-to-follow structure”* (RVS604).

Thanks to this summary activity, they also activated and exercised many metacognitive strategies to facilitate their second listening. For instance, this activity helped them to notice deficiencies in their previous interpretation of the listening content: *“In the first listening, as I lost my concentration in the final part of the video I did not understand much about the second experiment that the presenter carried out in the ocean”* (RVS603) or *“I have missed the information about different types of sea animals that can produce light in my first listening”* (RVS604). Interestingly, they all used these deficiencies to set the goals or foci for their second listening: *“In my summary, I missed one important example of a sea animal that can make light, which then directed my attention to this information in my second listening and then I noted it down here in red ink”* (RVS601) or *“I hesitated here [there was a long pause in the recording] as I missed the information about the purpose of this experiment. In the second time, I understood more about this purpose that was to observe how sea*

animals reacted to the optical lure” (RVS603). Two learners even used what they had summarized as an advance organizer to aid their second listening: *“I used the main ideas in my summary as a checklist to confirm my previous understanding of the video content, add further information, or correct my misunderstanding”* (RVS603) or *“I used the main ideas in my summary as signposts to aid my text comprehension in my second listening”* (RVS602).

According to these learners, it was the combined effect of these cognitive and metacognitive processes that helped to enhance their text comprehension. With this better text comprehension, they were more able to deploy available contextual clues for their interpretation about the meaning of novel words in the input material.

In addition, these learners also reported different benefits of this summary activity as an output task for their incidental vocabulary uptake. First, this activity required them to self-evaluate their knowledge of newly met words and decide whether or not to use these words in their summaries: *“I noted down here the word “plange” (an incorrect approximation of “plunge”) as it seems to be important to understand the final part of the talk. However, I have not learnt this word yet so I did not use it and also removed the idea related to it from my summary”* (RVS601) or *“I noted down this word (“bio-luminate” – an incorrect approximation of “bioluminescence”) because it is the key word in the video. However, I just caught bio-something in the first listening so I did not use it and used “making light” instead in my summary”* (RVS603). In these instances, both learners were found to obtain the precise meaning of these words on the word-meaning recall post-test, but neither got the precise form of these items on the word-form recall post-test.

Second, through this activity, they further noticed gaps in their lexical resources and then directed their attention to relevant lexis in the second listening: *“I know this word [bioluminescence] is a key word in the talk. I can guess its meaning, but I cannot catch its sound in my first listening and I do not use it in my summary. In the second time, I pay attention to it and note it down here how it might sound like”* (RVS601) or *“In the summary, I used “a machine” rather than “optical lure”. It is because I just caught its sound in my first listening, but not yet its meaning. So I did not use it. In the second time, I tried to guess its meaning again. It is a device to test how sea animals react to its light, artificial light”* (RVS603). For these instances, again both learners obtained the precise meaning of the target words on the word-meaning recall post-test. Only one (RVS604) had the precise form of the

target words on the word-form recall post-test. RVS604 gave “*biolation*” as the spoken form of “*bioluminescence*”.

Finally, this activity created opportunities for them to use newly met words and therefore foster their retention of these lexical items in the post-test: “*I have never seen the word “lure” or “jelly” before. I don’t know the meaning but I still use them in my summary and then guess their meaning in my second viewing time*” (RVS601) or “*I never see or use the word “bioluminescence” before but I guess its meaning by its prefix “bio-” as something related to living things and “-lumi-” as something related to light so I used this word in my summary and remembered it*” (RVS602). In fact, for any newly met words that the learners did use in their summaries, they obtained the precise form and meaning of these lexical items in the post-test.

It should be also noted that their use of the target words was all grammatically correct, semantically appropriate, and in syntagmatic contexts that distinctively differed from the original phrasings in the videos: “*The second experiment uses an optical lure. This is a small machine put under the sea. It can produce light as the way sea animals produce light. A lot of jellyfish came and appeared interested in the lure. And a shark also came and tried to eat the lure.*” (RVS601) (cf. “This is an optical lure that I have used. We call it the electronic jellyfish”) or “*The speaker is trying to spark some interest in one phenomenon which is bioluminescence or the way sea animals make light – something she is really enchanted by.*” (RVS603) (cf. “So it is a little appreciated fact that most of the animals in our ocean make light. I have spent most of my career studying this phenomenon called bioluminescence. I study it because I think understanding it is critical to understanding life in the ocean where most bioluminescence occurs.”).

Altogether, the learners in this learning condition reported that the summary activity helped them to develop their lexical knowledge in two different ways. First, it prompted them to carry out different cognitive activities to reprocess perceived information from their first listening and different metacognitive processes to facilitate their second listening. It was the combined effect of these cognitive and metacognitive processes that helped to enhance their text comprehension. This better text comprehension, in turn, put them in a better position to make use of available contextual clues for their word-meaning inference. Second, as an output task, this summary activity gave them opportunities to self-evaluate their existing knowledge of newly met words and then decide whether to use these words in their

summaries. When they did use these words in their summaries, this use helped them to create a durable memory of these lexical items. This activity also prompted them to notice gaps in their vocabulary repertoire and then direct their attention to relevant lexis in the second listening. Incidental vocabulary acquisition did happen on this premise.

Benefits of the Interactional Summary Activity in the NVS Condition

Quantitative results

Table 29 presents the learners' raw scores on the text comprehension measure and the difference between their pre-test and immediate post-test scores on the vocabulary measures:

Table 29: Text comprehension and Vocabulary gain in the NVS condition

Learners	Text comprehension	Vocabulary Gain	
		Form Recall	Meaning Recall
NVS605	10	3/37	
NVS606	10	3/35	
NVS607	12		6/31
NVS608	13		5/26

On the text comprehension measure, these learners' total scores ranged between 10 (67%: NVS605, NVS606) and 13 (87%: NVS608) out of the maximum score of 15 points. NVS605 and NVS606 made the same gain of 3 out of 37 (8%: *lure*, *plankton*, and *plunge*) and 35 (9%: *camouflage*, *lure*, and *plankton*) target words on the word-form recall test, respectively. Meanwhile, NVS607 and NVS608 obtained a gain of 6 (19%: *bioluminescence*, *joystick*, *junky*, *lure*, *plankton*, and *plunge*) and 5 (19%: *bait*, *enchanted*, *lure*, *outrageous*, and *plunge*) out of 31 and 26 target words on the word-meaning recall test, respectively. Together, the four learners in this learning condition also achieved a considerable degree of text comprehension and vocabulary gain from their assigned viewing procedure.

Qualitative results

Like those in the RVS condition above, the learners in this learning condition also reported that the summary activity prompted them to reprocess perceived information from their previous listening, which, in turn, enhanced their text comprehension. Specifically, it helped them to differentiate the main ideas of the videos from the supporting ideas and factual information (NVS605, NVS607, and NVS608), organize perceived information into a coherent structure (NVS607 and NVS608), notice further deficiencies in their previous

understanding of the input content (NVS605, NVS606, NVS607, and NVS608), and also reinforce their memorization of the listening content (NVS607 and NVS608).

In the present summary activity, the learners had the opportunity to interact with their interlocutors to aid their understanding of the given summaries. For each pair of learners, I found two episodes where they negotiated the meaning of the video content. In both episodes, the learners sought for content clarification. Below is one of these examples:

NVS605: *So you mean they [scientists] used Titanic to explore the ocean?*

NVS606: *No. I mean they dreamt to build up a special technology to explore Titanic.*

NVS605: *What kind of technology is this?*

NVS606: *Sorry. I didn't catch it.*

It is clear from this example that when NVS605 asked for content clarification, NVS606 tried to elaborate her idea, and further noticed shortcomings in her previous understanding of the input text. In the follow-up interview, NVS606 also added that whenever she got such a question from her interlocutor, she had to “*reprocess the information again, reorganize it in a way to get the message across*” and “*this also improves [her] text comprehension*”. NVS607 suggested another benefit of such interaction: “*Through this interaction, I not only seek for content clarification but also note this content down on paper in order to check it out in my upcoming viewing of the video*”.

In this summary activity, the learners were given a summary of the video content before they watched the video(s) themselves. They all confirmed that such a summary facilitated their actual listening comprehension in various ways. First, NVS606 and NVS607 reported that this summary gave them the topic of the upcoming video(s), which therefore helped to effectively activate their top-down processing: “*Through her summary, I know the topic of the upcoming video*” (NVS606) or “*This summary is very helpful for my actual video watching as I know the topic and thus I can predict its content*” (NVS607). NVS605 and NVS606 used the sequence of information given in this summary as an advance organizer to foster their actual listening: “*Through her summary, I can catch some key words which I use as signposts to facilitate my actual listening*” (NVS605) or “*I use the key events in her summary to predict what is coming next in my actual listening*” (NVS606). Interestingly, based on the given information in this summary, all learners formed hypotheses about the upcoming content. Therefore, in their actual listening session, they merely needed to

“confirm what has been said” (NVS607), *“add further detailed information to the key point in the summary”* (NVS608), *“reject this information when it is not mentioned or incorrect”* (NVS605). Thanks to these benefits, they found their actual listening *“much easier”* (NVS608) and thus they *“enjoyed listening”* rather than made effort to take in the video content (NVS607).

The report above clearly indicates that, on the one hand, giving a summary of the video content after viewing prompted the learners to carry out different cognitive activities to reprocess perceived information from their previous listening which therefore enhanced their comprehension and retention of the listening content. On the other hand, receiving a summary of the video content before viewing helped the learners to implement such metacognitive strategies as using the given topic, key information, and sequence of this information to activate their top-down processing, which, in turn, fostered their actual listening comprehension.

However, according to these learners, it was receiving a summary of the video content before viewing rather than giving a summary of the video content after viewing that helped them with their incidental vocabulary acquisition. This was because this summary reduced the cognitive load invested for the purpose of text comprehension and they thereby had more mental resources to deal with new lexical items in the input materials. This finding is consistent with what I found in Study 5.

As an output task, this summary activity provided the learners with many other benefits for incidental vocabulary uptake. Like the output task in the case of the RVS condition above, this summary activity also prompted the learners to self-assess their knowledge of newly met words and decide whether or not to use these words in their summaries (NVS605, NVS606, and NVS607), notice further lexical gaps in their inter-language system (NVS605, NVS606, NVS607, and NVS608), and use newly met words in their summaries and therefore consolidate their knowledge of these lexical items (NVS605, NVS606, NVS607, and NVS608).

In this summary activity, the learners were allowed to negotiate the form and meaning of any unfamiliar words including the target words with their interlocutors. I found altogether three instances where the learners negotiated the form and/or meaning of unfamiliar words (*burglar alarm*, *joystick*, and *lure*). Two of these instances were related to the target words in

this study (*lure* and *joystick*). In these cases, both word givers and receivers remembered these lexical items in the post-test. Below is one example:

RVS608: *Their dream was to help people to learn about Titanic right from home by just using joystick and headphone.*

RVS609: *What was the jointstick?*

RVS608: *A joystick is just like the equipment that you use to control like...Have you played games?*

RVS609: *What kind of games [giggle]?*

RVS608: *Like Xbox.*

RVS609: *No.*

RVS608: *OK. Can I just spell it and you look for it later?*

RVS609: *Can you explain it in Vietnamese?*

RVS608: *Thiết bị điều khiển khi em chơi game* [Vietnamese: a remote control you use to play games]

RVS609: *Yeah. Thanks.*

After this interaction, RVS608 jotted down the word “*jointstick*” rather than “*joystick*” on her note sheet (see Appendix I). A double check with their test scores showed that while RVS608 already knew this word prior to the experiment, RVS609 did not know it yet. However, RVS608 got a precise meaning of this word in her word-meaning recall post-test thanks to the negotiation episode above. Thus, when the learners were allowed to negotiate the meaning of unfamiliar words in this activity, there was evidence that they could increase their knowledge of these lexical items. RVS609 also confirmed in the follow-up interview that she picked up the meaning of this word from her interlocutor’s word-meaning explanation.

Finally, listening to a summary of the input content provided opportunities for the learners to recycle and even expand their vocabulary knowledge: “*When I listened to her summary, I recognized the word “defence” that I never used before and I also learnt a new word “enemy” as something similar to “predator” and I actually used “enemy” in my summary*” (RVS606).

Taken together, giving a summary of the video content after viewing prompted the learners to carry out different cognitive activities to reprocess their perceived information from the previous listening and therefore fostered their text comprehension. Receiving a summary of the video content before viewing helped them to implement different metacognitive strategies to facilitate their upcoming listening. However, according to these learners, it was the latter rather than the former that aided their incidental vocabulary acquisition. As an interactional output task, this summary activity gave them many other benefits for incidental vocabulary uptake. First, it helped them to self-assess their current knowledge of newly met words and decide whether to use these words in their summaries. In case they did use these words in their summaries, this use not only fostered their memorization of these lexical items but also created opportunities for negotiation and then uptake of the word meaning.

An Overall Summary of the Research Results

The results from this research project are summarized in Table 30 below:

Table 30: Summary of the research results				
Conditions	Form gain		Meaning gain	
	Effect size*	Mean gain**	Effect size	Mean gain
OV	1.47	0.5	0.84	0.5
RV	1.53	1	1.66	1
Comparison of vocabulary gain	Comparison		RV > OV	
	Explanatory factors		Text comprehension	
	Further explanation		The second listening helped to improve text comprehension, which, in turn, fostered incidental vocabulary acquisition.	
RV	1.53	1	1.66	1
RVS	1.55	2	1.93	2
Comparison of vocabulary gain	Comparison		RVS > RV	
	Explanatory factors		<ul style="list-style-type: none"> - Text comprehension - How the learners dealt with the target words in the summary activity 	
	Further explanation		<ul style="list-style-type: none"> - The combined effect of cognitive and metacognitive processes that the learners used to deal with the summary activity helped to enhance text comprehension, which, in turn, fostered incidental vocabulary acquisition. - As an output task, this activity gave the learners 	

			opportunities to use the target words in their summaries and then created a durable memory of these words. It also helped them to notice gaps/holes in their lexical resources and then prompted them to fill these gaps/holes in the second listening.	
OV	1.52	0.5	0.86	0.25
NV	1.58	0.5	1.00	0.5
Comparison of vocabulary gain	Comparison		NV > OV	
	Explanatory factors		Text comprehension	
	Further explanation		Receiving the two shorter videos of virtually the same topic beforehand helped to improve comprehension of the longest video, which, in turn, fostered incidental vocabulary gain from this video.	
NV	1.14	1	1.02	1
NVS	1.77	2	2.77	2
Comparison of vocabulary gain	Comparison		NVS > NV	
	Explanatory factors		Text comprehension	
	Further explanation		Receiving a summary of the video content before viewing helped to improve text comprehension, which, in turn, fostered incidental vocabulary acquisition.	
*Cohen's d values; **Number of target words (as measured on the immediate post-test)				

It is clear from this table that all interventions suggested in this research project indeed brought about statistically significant vocabulary gains and the size of vocabulary gains differed from one intervention to another. However, it also needs to be acknowledged that if we view these gains from the number of novel words actually picked up from each intervention these gains turn out to be relatively modest. As hypothesized, text comprehension significantly predicted the size of vocabulary gains both within and between the interventions. Additionally, how the learners dealt with the target words in the summary tasks in the RVS intervention and the NVS intervention was also found to affect the amount of their incidental vocabulary gain and retention.

CHAPTER 4: DISCUSSION AND IMPLICATIONS

In the first section of this chapter, I shall discuss the key factors that were found to account for the vocabulary gain in each treatment condition and the difference in the size of vocabulary gains between these conditions. Subsequently, I will relate these factors with seven dominant theories, Ausubel's Advance Organizer (1978), Hulstijn and Laufer's Involvement Load Hypothesis (2001), Krashen's Comprehensible Input Hypothesis (1985), Long's Interaction Hypothesis (1996), Nation's Vocabulary Generation (2013), Swain's Output Hypothesis (2005), and Wittrock's Model of Generative Teaching of Comprehension (1991), together with the major research findings in the fields of L2 listening comprehension and incidental vocabulary acquisition. The outcome of this discussion will help me to draw out some implications for instructional practice and future research in the second section.

Discussion

Study 1: The Effect of Viewing a TED Talks Video Once (OV)

The results from this study indicate that L2 learners were able to pick up some new lexical knowledge from a single viewing of a TED Talks video. They obtained a mean gain of 0.57 (4.39%) and 0.28 (2.15%) out of 13 target words on the word-form recall measure and the word-meaning recall measure, respectively. In comparison with a control group, this vocabulary gain reached the required significance level in the case of the word-form gain (with large effect sizes of 1.52 and 1.51 for acquisition and retention, respectively), but it fell short of significance in the case of the word-meaning gain (but nonetheless showing a large effect size of 0.84 for acquisition and a medium effect size of 0.70 for retention). This finding appears to support the claim made by Coxhead and Walls (2012), Floyd and Jeschull (2012), and Takaesu (2013) that TED Talks have the potential to foster incidental L2 vocabulary uptake.

It might be interesting to compare the effect sizes of the treatment condition in this study with those reported in previous research. Rodgers (2013), for example, found that the L2 learners in his treatment group ($N = 178$) obtained a mean gain of 6.36 (10.60%) out of 60 target words after viewing 13 episodes of a TV program over a period of 13 weeks (one episode per week). However, those in his control group, who did not view any of these episodes throughout the same semester ($N = 73$), were also found to make a mean gain of 5.32 (8.87%) target words. In other words, the learners in his treatment group merely picked

up only 1.04 (1.73%) target words more than those in the control group. The effect size of the treatment condition in this study was indeed relatively small: $d = 0.41$ - much smaller than in my present study. Vidal (2011) also measured incidental L2 vocabulary gain, but from viewing three video-recorded lectures over a period of four weeks. After the experiment, the learners in her treatment group ($N = 112$) were found to obtain a mean score of 28.94 (SD 22.35) out of the maximum score of 180 points on the vocabulary measure. Those in the control group, who did not view any of these lectures during the entire studied period ($N = 38$), achieved a mean score of 2.24 (SD 22.35). The effect size of the treatment condition in this study was 1.38, which was slightly smaller than those on word-form gain, but noticeably larger than those on word-meaning gain in my present study. Together, these comparisons suggest that the lexical uptake in my present study was relatively limited, especially in terms of the number of the target words being acquired. This finding seems to reflect the slow nature of incidental vocabulary acquisition proposed by Nagy, Herman, and Anderson (1985). It also confirms the challenges of incidental L2 vocabulary acquisition from listening input (van Zeeland, 2013), and this includes TED Talks (Coxhead & Walls, 2012).

Following Krashen's Comprehensible Input Hypothesis (1985), this small vocabulary gain can first be attributed to the limited text comprehension that the learners obtained from viewing the video only once. On the text comprehension measure, the learners in this study indeed achieved a mean score of only 5.13 (SD 1.43) out of the maximum score of 15 points (34.20%). If one accepts a score of 7.50 points (50%) as a threshold indicating an adequate understanding of the video content, only one learner in this study passed this level. At a stricter threshold of 10.50 points (70%), none reached this level. This limited text comprehension may be caused by the challenging nature of the input material. The video used in the present study is authentic in a sense that it was produced without L2 learners in mind (Coxhead & Walls, 2012), and then brought into the experiment without any modification. Such characteristics of authentic listening materials as natural and fast speed of delivery might be challenging to the learners (Vandergrift, 2004). It is also worth noting in this regard that most previous research has used either language-instructional materials like graded readers (e.g., Brown et al., 2008; Webb & Chang, 2012) or self-constructed materials for research purposes (e.g., van Zeeland & Schmitt, 2013; Vidal, 2011) as listening input. In addition, having the learners view the video only once may constitute another impediment to their text comprehension (Berne, 1995; Cervantes & Gainer, 1992; Chang & Read, 2006; Iimura, 2007; Lund, 1991; Saika, 2009). When the learners struggled to understand the video

content, most of their working memory capacity might have been invested for the purpose of text comprehension and the remaining mental resources would be insufficient to deal with new lexical items in input materials (Boers et al., 2007). Although the results from the Pearson Correlation calculation show that the positive coefficients between lexical uptake and text comprehension fell short of significance, this might be due to the small sample size and the very small range of vocabulary gains in the present study.

Following Webb (2014), another plausible explanation for this modest vocabulary gain is the low frequency of word occurrences in the input material. As the video in this study was not modified, word repetition could not be manipulated. As a result, all but one target word (*bioluminescence*) in the present study occur only one or two times in the video, which is much lower than 10+ or 20+ times suggested respectively by van Zeeland and Schmitt (2013a, p.621) or Brown et al. (2008, p.153) as necessary for incidental vocabulary acquisition from listening input to occur.

According to Neuman and Koskinen (1992), Sawada (2009), Sydorenko (2010), and Vidal (2003, 2011), another influential factor in incidental L2 vocabulary acquisition from videotexts is the presence of both verbal and pictorial clues for word-meaning inference. However, this factor seems not to have made much difference in the present study. In fact, there are many target words in the given video such as “*bioluminescence*” or “*plunge*”, which are surrounded by useful verbal and pictorial clues for word-meaning interpretation. However, the learners’ limited text comprehension may have restricted their ability to deploy these clues for successful word-meaning inference.

In addition, TED Talks are often found to contain many technical words, whose concepts might be unfamiliar to the learners (Coxhead & Walls, 2012). The talk used in the present study is no exception. “*Bioluminescence*”, “*optical lure*”, or “*trawling*” are just a few examples of such technical concepts. Therefore, the learners in this experiment may have needed to establish these new concepts in mind first rather than directly map known concepts to L2 forms (Vidal, 2011). The presence of these technical concepts may have impeded their listening comprehension and incidental vocabulary acquisition at the same time.

In the present study, the learners’ acquired lexical knowledge proved to be relatively resistant to decay for at least two weeks after exposure. This finding is different from what Sawada (2009) or Vidal (2003, 2011) have found in their studies. However, this might be attributed to the floor effect of such a modest vocabulary gain in this experiment.

Summing up, this study provides empirical evidence for incidental L2 vocabulary uptake from viewing a TED Talks video only once. This therefore seems to support the claim made by Coxhead and Walls (2012), Floyd and Jeschull (2012), and Takaesu (2013) about the benefits of TED Talks for incidental vocabulary acquisition. However, the size of vocabulary gain was very modest. This confirms the challenges of incidental vocabulary acquisition from listening input (van Zeeland, 2013), including TED Talks (Coxhead & Walls, 2012). The learners' poor text comprehension, the presence of many technical words with unfamiliar concepts, and the limited frequency of word occurrences in the input material also help to explain this modest vocabulary gain. These findings suggest that incidental L2 vocabulary acquisition from TED Talks should not be taken for granted or, in other words, that teachers and course designers might need to apply some pedagogical intervention to increase the chances of vocabulary uptake from this source of input.

Study 2: Adding another Viewing Time (RV)

One pedagogical intervention I suggested in this research project was to give L2 learners another viewing of the same TED Talks video (i.e., repeated viewing, or RV). This suggestion was experimented in Study 2 with another group of EFL learners. The results from this study show that these learners, on average, picked up one out of the 17 (5.88%) and 18 (5.56%) target words on the word-form recall measure and the word-meaning recall measure, respectively. Albeit such a limited number of target words gained in the treatment condition, this condition produced a large effect size of 1.53 on word-form acquisition and of 1.51 on the word-form retention. The effect sizes in the case of word-meaning acquisition and retention were 1.66 and 1.11, respectively.

It is also worth comparing these effect sizes with those documented in previous research. Rodgers (2013) and Vidal (2011), for example, found that their treatment conditions had effect sizes of 0.41 and 1.38 on vocabulary gain, respectively. It is obvious that these effect sizes were noticeably smaller than those in my present study. In addition, Study 6 indicates that the RV condition also created larger effect sizes on vocabulary gain than the OV condition, especially in the case of word-meaning gain. This better lexical uptake can be attributed to the benefits of the second viewing time, as discussed below.

First, another viewing of the same TED Talks video helped the learners in the RV group to achieve better text comprehension than those in the OV (i.e., One Viewing) group. On the text comprehension measure, the former were found to get a mean score of 7.25 (*SD*

1.24) out of the maximum score of 15 points (48.33%) while the latter only obtained a mean score of 5.13 points (*SD* 1.43) (34.20%). Study 6 indeed shows that the RV condition led to significantly better text comprehension than the OV condition. This finding echoes what Berne (1995), Cervantes and Gainer (1992), Chang and Read (2006), Iimura (2007), Lund (1991), and Saika (2009) have found about the positive effect of repeated listening on L2 listening comprehension.

In the case of the present study, repeated viewing might have benefited the learners' text comprehension both affectively and cognitively. As the learners were informed beforehand that they were going to watch the video twice rather than once, they might have felt less anxious and therefore more engaged with the listening process (Berne, 1995). As discussed in the Literature Review Chapter, the limited capacity of human working memory (Cowan, 1995; Robinson, 2003; Schmidt, 1990) and the real-time nature of listening (Buck, 2001) often make it challenging for L2 learners to hold in memory what they have heard while at the same time struggling to keep up with incoming information or, conversely, to focus on the incoming information while still struggling to make sense of what preceded (Goh, 2000). Thus, having another viewing of the same video gave the learners in the RV condition the opportunity to verify their previous interpretation of the video content and also to add information they missed in their first viewing.

Following Krashen's Comprehensible Input Hypothesis (1985), a better degree of text comprehension puts learners in a better position to make use of available contextual clues for their word-meaning interpretation. As the learners in the RV condition were found to achieve better text comprehension than those in the OV condition, the former were also expected to make a larger vocabulary gain than the latter. This expectation is indeed supported by a positive and moderate correlation between the text comprehension scores and the vocabulary gain scores obtained by the learners in these conditions. The results from the Multiple Regression analyses in Study 6 provide further supporting evidence for this expectation. This finding is therefore consistent with what Sawada (2009) and Vidal (2003, 2011) have found in their experiments. Together, it might be concluded that a sufficient level of text comprehension in the RV condition created a favourable (but not necessarily sufficient) condition for incidental vocabulary acquisition to occur, especially during the second viewing.

Another factor that was also found to facilitate incidental vocabulary acquisition in this study is the higher frequency of word occurrences in the input material. Having the learners view the same video twice doubled the frequency at which they encountered the target words in the input material as compared to having them view this video only once. A significant, positive, and moderate coefficient was indeed found for the association between word repetition and word-meaning gain in this study ($N = 18$). This association in the case of word-form gain was also positive and moderate, but fell short of the required significance level, which might be attributed to the smaller sample size on this measure ($N = 14$). This finding is in line with what Brown et al. (2008), Horst et al. (1998), Rodgers (2013), van Zeeland and Schmitt (2013) and Vidal (2003, 2011) have observed in their experiments. This also supports Webb's (2014) claim about the important role of word repetition in incidental vocabulary acquisition.

A pertinent question here is whether it was their improved text comprehension or the higher frequency of their word encounters that gave the RV group an advantage over the OV group for incidental vocabulary acquisition. The Multiple Regression Analyses in Study 6 reveal that the larger vocabulary gain made by the former was better explained by their better text comprehension than by the higher frequency at which they encountered the target words in the listening input. It should be also noted that, unlike the case in most previous research, the target words in the present study were reencountered in identical contexts. On the one hand, this type of word reencounters allowed the learners to make use of their previous understanding of the video content to infer the meaning of the target words. On the other hand, it also entailed much fewer opportunities for the learners to fine-tune their inferences than if they had been exposed to an accumulation of contextual clues (R. Ellis, 1999).

In Study 1, the learners were hindered in their use of available contextual clues for word-meaning interpretation by their limited text comprehension. In the present study, as the text comprehension was noticeably enhanced, the evidence for the association between the usefulness of verbal clues in the video for word-meaning inference on the one hand and the size of word-meaning acquisition ($r = .41, p = .09$) and retention ($r = .45, p = .06$) on the other hand also became clearer, even though the required significance level was yet to be reached. Again, this may be due to the small sample size in the present experiment.

Altogether, viewing a TED Talks video twice indeed led to better lexical uptake than viewing the same video only once. This better vocabulary gain can be directly attributed to

the benefits of the second viewing, where the learners had the opportunity to reencounter the target words in identical contexts, and then made use of their previous understanding of the listening content to infer the meaning of these lexical items. Thus, the findings of this study support Krashen's Comprehensible Input Hypothesis (1985).

Study 3: Adding an Output Task between the Two Viewing Times (RVS)

The aim of this study was to measure the size of vocabulary gain from viewing a TED Talks video, making an oral summary of the content, and then viewing this video again. The results from this study show that L2 learners were indeed able to make a statistically significant vocabulary gain from this learning condition. On average, these learners picked up two out of the 17 (11.77%) and 18 (11.11%) target words on the word-form recall measure and the word-meaning recall measure, respectively. It should be also noted that this vocabulary gain was resistant to decay until even two weeks after the experiment. The effect sizes that this learning condition created on word-form acquisition and retention were large with $d = 1.55$ and 1.63 , respectively. These effect sizes in the case of word-meaning acquisition and retention were even larger with $d = 1.93$ and 2.22 .

As compared with the effect size of 0.41 in the experiment by Rodgers (2013) and 1.38 in that by Vidal (2011), the effect sizes of this learning condition were considerably larger, especially in the case of word-meaning gain. The results from Study 6 also indicate that the effect sizes of this learning condition even surpassed those in the RV condition. This better vocabulary gain can be attributed to the benefits of the summary activity, as discussed below.

First, this summary activity helped the learners to achieve better text comprehension. On the text comprehension measure, the learners in the RVS condition obtained a mean score of 9.32 ($SD\ 1.18$) out of the maximum score of 15 points (62.13%). Recall that those in the RV condition only achieved a mean score of 7.25 ($SD\ 1.24$) (48.33%). Study 6 shows that the former had markedly better text comprehension than the latter. This finding appears to support Wittrock's claim about the benefit of using a summary activity to foster text comprehension (1991).

According to Krashen's Comprehensible Input Hypothesis (1985), a better level of text comprehension makes learners more able to deploy available contextual clues for their word-meaning inference and therefore helps enlarge the size of their vocabulary gain. As the

learners in the RVS condition were found to obtain better text comprehension than those in the RV condition, the former were also expected to make a larger vocabulary gain than the latter. This expectation is again supported by a positive and moderate correlation between the text comprehension scores and the vocabulary gain scores obtained by these groups.

However, the Multiple Regression analyses in Study 6 also suggest that there was another factor which could explain the difference in the vocabulary gains between these groups. In what follows, I shall focus on this factor.

As an output task, the summary activity in the present study was found to stimulate incidental vocabulary acquisition in line with three major functions in Swain's Output Hypothesis (2005): hypothesis testing function, fluency development function, and noticing function. First, eleven out of the 32 learners in this experiment were found to use at least one target word in their summaries. It should be noted that the syntagmatic contexts in which the target words were used all differed from the original phrasings in the given input. According to Swain (2005), in order to use a newly met word in a novel context, learners need to form a hypothesis about how this word should work in that particular context. When interactional feedback is available, learners are able to use this feedback either to confirm, modify or reject the hypothesis, and then fine-tune their lexical knowledge. Although this feedback was, by nature, unavailable in this learning condition, the learners could still verify their hypotheses by comparing their use of target words in their oral summary with the way these words were used in the TED Talk during their second viewing.

Following the fluency development function comprised in Swain's Output Hypothesis, when learners attempt to use a newly met word in an output task, they first need to retrieve it from their memory. Successful retrieval of this lexical item from the memory is believed to strengthen the memory path to that item, which, in turn, enhances learners' automaticity in the word recognition and recall in the future. In addition, Joe (1998) also found from her experiment that a novel use of a newly met word in a text-based output task created a more durable memory of this word than a verbatim replication. The findings in my present study seem to be consistent with what Joe (1998) reported in her study and Nation (2013) claimed about the benefit of generative processing for lexical retention. Regarding four target words that were actually used in the summary task, the learners did not know these words prior to the experiment (as evidenced by their pre-test performance), but they did use these lexical items in new syntagmatic contexts (as compared to the original phrasings in the video), and then successfully recalled these items even two weeks after their exposure.

Swain further suggested as part of her Output Hypothesis that an output task helps learners to notice gaps in their lexical resources, which, in turn, prompts them to fill these gaps when they are exposed to L2 input again. In the case of the present study, I found 36 instances of conspicuously long (more than 2.00 seconds) pauses where a target word could have been appropriate. All the pauses occurred in the middle of a clause, suggesting that the learners attempted to recall the target words from their memory, but failed to do so (Tavakoli, 2011). It is interesting to note that, out of these 36 retrieval failure instances, 13 were followed by successful word recall in both post-tests. It is conceivable that the retrieval failures prompted the learners to direct their attention towards relevant lexis in their second viewing. Incidental vocabulary acquisition may have happened on this premise. This therefore appears to support Swain's noticing function.

It should be, however, noted that the nature of the output task in my present study is distinctly different from those in previous research. L2 learners in most previous studies were first provided with the meaning of the target words, and then required to use these words in their output tasks (e.g., de la Fuente, 2002; Ellis & He, 1999). Clearly, one aim of these tasks was to help L2 learners to recycle their recently learnt words. Put differently, these tasks were either *focus on form* (Long, 1988, 1991; Long & Crookes, 1992) or *planned focus on form* (Doughty & Williams, 1998; Harley, 1998; Swain, 1998). Meanwhile, the output task in my present study prompted the learners to deal with the listening content in order to enhance their text comprehension. In other words, this output task focused on meaning and only provided *incidental focus on form* (Ellis, 2001). When these learners needed to use any of the target words in their summaries, they must have succeeded in deciphering the meaning of these words from the surrounding context during their preceding listening. In most cases, however, the learners in the present study were not confident enough about the amount of knowledge that they had acquired for these target words. As a result, they avoided using these words, and then resorted to paraphrasing these lexical items in their summaries. This might explain why the effect of this output task on vocabulary gain was not as large as those of the output tasks in previous research. It is nevertheless clear from the discussion above that such a meaning-focused output task also benefited incidental vocabulary acquisition, but mainly through enhancing the learners' text comprehension and, consequently, fostering their lexical uptake in the second viewing.

Although the learners in this learning condition were totally free in their choice of vocabulary use in the output task, the nature of the task – summarizing the listening content –

tended to fasten their focus on the key words in the input material. Thus, it came as no surprise that the top three target words learnt from their preceding listening and used in their summaries were also three key concepts in the video: “*bioluminescence*”, “*bait*”, and “*lure*”. This suggests that the nature of the announced text-based output task may have affected the way the learners processed the given input text.

It should be noted that the learners in this learning condition not only made a larger word-meaning gain but also obtained better a word-form gain than those in the RV condition. This better word-form gain might be due to the effect of informing the learners beforehand about the presence of the summary task after their first viewing. This announcement might also have prompted them to process the listening input in a different way. To be more specific, if the learners in the RV condition tended to focus more on the meaning of the target words as they were required to answer the content questions afterwards, those in the RVS condition needed to process both the form and the meaning of these lexical items as they knew in advance that they might find using these items useful to summarize the listening content after their first viewing. This finding is consistent with what Foster and Skehan (2013) and Stahl and Clark (1987) have found about the positive effect of anticipating a post-task output activity on focus on form and incidental vocabulary acquisition.

As the learners in the present experiment were exposed to the video twice, their incidental vocabulary acquisition was also found to benefit from their reencounters with the target words in the second viewing. In fact, the frequency of word occurrences in the input material was found to be positively and moderately correlated with the size of their word-meaning gain. This finding is consistent with what Brown et al. (2008), Horst et al. (1998), Rodgers (2013), van Zeeland and Schmitt (2013) and Vidal (2003, 2011) have found in their experiments. This again also support Webb’s (2014) claim about the important role of word repetition in incidental vocabulary acquisition.

In Study 1 and 2, the evidence about the association between the usefulness of verbal clues in the video for word-meaning inference and the size of vocabulary gain was not clear. However, the size of lexical uptake in the present study was substantially larger than those in Study 1 and 2 and therefore this association became more clearly evidenced. In fact, the usefulness of verbal clues in the video for word-meaning interpretation was found to be positively and moderately correlated with both word-form and word-meaning gain. This

finding is similar to what Neuman and Koskinen (1992), Sawada (2009), Sydorenko (2010), and Vidal (2003, 2011) have reported in their studies.

To summarize, the results from this study indicate that the learners in the RVS condition indeed made a statistically significant vocabulary gain from their assigned viewing procedure. This vocabulary gain was even larger than that in the RV condition thanks to two benefits of the summary activity. First, this activity helped the learners to enhance their text comprehension. This improved text comprehension enabled them to make better use of available contextual clues for word-meaning inference, especially in their second viewing. This therefore supports both Wittrock's Model of Generative Teaching of Comprehension (1991) and Krashen's Comprehensible Input Hypothesis (1985). Second, this activity also provided opportunities for the learners to (a) form and test their hypotheses about how the target words should work in new contexts, (b) use the words in these contexts and foster their memorization of these lexical items, and (c) notice further gaps in their lexical resources and then direct their attention to relevant lexis in the second viewing. All these opportunities might have helped them to enlarge the size of their vocabulary gain. In other words, these findings are consistent with four major functions that Nation (2013) and Swain (2005) attributed as the benefits of an output task for vocabulary learning: generative processing, hypothesis testing, fluency development, and noticing. It should be also noted that, in this study, the target words which were necessary for text comprehension and/or occurred more often in the input text were more likely to be acquired. This therefore supports the claim made by R. Ellis (1991) and Webb (2014) about the saliency of key words and word repetition in incidental vocabulary acquisition.

Study 4: Adding Two Other Videos about Similar Topics (NV)

This study examined the degree to which L2 learners were able to pick up the form and meaning of novel words from viewing three TED Talks videos on similar topics, including the video used in Study 1, 2, and 3. On average, the learners who viewed these videos learnt one out of 25 (4.00%) and 29 (3.45%) target words on the word-form recall measure and the word-meaning recall measure, respectively. In comparison with the control group (who did not watch the videos) this learning condition created a large effect size of 1.14 and 1.32 on word-form acquisition and retention, respectively. These effect sizes in the case of the word-meaning acquisition and retention were also large with $d = 1.02$ and 1.09. These effect sizes are much larger than that of 0.41 in the study by Rodgers (2013), but

slightly smaller than that of 1.38 in the study by Vidal (2011). However, it should be also noted that the number of the target words learnt in this study was considerably smaller than the number of target words in the study by Vidal (2011).

On the dataset of the 20 target words from the longest video, the learners in this condition were also found to obtain a larger vocabulary gain than those in the OV condition, especially in the case of word-meaning gain. The effect sizes that the NV condition produced on word-meaning acquisition and retention were large with the same d value of 1.00. Meanwhile, these effect sizes in the case of the OV condition were noticeably smaller with $d = 0.86$ and 0.71 , respectively. In what follows, I shall look at the factor that was found to account for this finding.

This larger word-meaning gain can be attributed to the fact the learners in the NV condition obtained better text comprehension for the longest video than those in the OV condition, as shown in Study 6. Out of the five comprehension questions regarding the content of the longest video, the NV group made a mean score of 2.25 (SD 1.11, 45%) whereas the OV group merely obtained a mean score of 1.23 (SD 0.99, 24.60%). This better text comprehension might result from the fact that the learners in the NV condition had the opportunity to view the two shorter videos of virtually the same topic as the longest one before they came to the longest video while those in the OV condition did not. In this way, the learners in the NV condition may gradually have familiarized themselves with the listening topics when they moved on from one video to another. Previous research has indicated a positive effect of topic familiarity on L2 listening comprehension (Leeser, 1998; Long, 1990; Sadighi & Zare, 2006; Schmidt-Rinehart, 1994; Zidong, 1998) and on incidental L2 vocabulary acquisition from reading (Pulido, 2016).

As reported in the Methodology Chapter, when the learners viewed these videos in the sequence from the shortest to the longest one as in the present study, they would considerably increase their lexical coverage for the last two videos in this video set. Previous research has also found a positive correlation between lexical coverage and L2 listening comprehension (Bonk, 2000; Stæhr, 2009). Thus, this finding supports Krashen's proposal of Narrow Listening (1996).

Again, following Krashen's Comprehensible Input Hypothesis (1985), this better text comprehension was hypothesized to put the learners in the NV condition in a better position to use available contextual clues for their word-meaning inference and therefore enlarge the

size of their vocabulary gain. This hypothesis was first supported by a positive and relatively strong correlation between the text comprehension scores and the vocabulary gain scores obtained by the learners in this experiment. The Multiple Regression Analyses in Study 6 provide further supporting evidence for this hypothesis. This finding is in line with what Sawada (2009) and Vidal (2003, 2011) have found in their experiments.

On the dataset of the 40 target words from all three videos, a positive and moderate correlation was also found for the relationship between the usefulness of verbal clues in the videos for word-meaning inference and the size of vocabulary gain in this condition. A closer look at the raw data shows that the top three target words with the most useful verbal clues for word-meaning interpretation (“*bioluminescence*”, “*puddle*”, and “*startling*”) were also the three target words for which the largest vocabulary gains were attested. This finding is in line with what Neuman and Koskinen (1992), Sawada (2009), Sydorenko (2010), and Vidal (2003, 2011) have found in their experiments.

According to Webb (2014), another factor that may account for the noticeable vocabulary gain in the NV condition is the frequency of word occurrences in the input materials. In this learning condition, word repetition was indeed found to be positively and moderately correlated with the size of word-form gain. This correlation in the case of word-meaning gain was also positive and moderate, but fell short of the required significance level. This might be attributed to the fact that the gain on the word-meaning recall measure was smaller than that on the word-form recall measure. This finding is consistent with that reported by Brown et al. (2008), Horst et al. (1998), Rodgers (2013), van Zeeland and Schmitt (2013), and Vidal (2003, 2011).

Taken together, having L2 learners view three TED Talks videos on similar topics also brought about a statistically significant vocabulary gain (albeit a relatively small gain in real terms). This vocabulary gain, especially in the case of word-meaning gain, was larger than only having them view the longest one out of these three videos. This better lexical uptake was indeed associated with the better text comprehension that the learners obtained thanks to the benefit of narrow viewing. Therefore, the findings from this study appear to be consistent with Krashen’s Comprehensible Input Hypothesis (1985) and his proposal of Narrow Listening (1996).

Study 5: Adding an Output Task into the Narrow Viewing Procedure (NVS)

The aim of this study was to measure the size of vocabulary gain from a learning condition where L2 learners were required to view either the first or the second half of the three-video set in Study 4, summarize the content to a classmate who had not viewed the same half yet in an interactional information-gap activity, and then move on with the other half. The learners in this learning condition were found to pick up two out of 25 and 29 target words on the word-form recall test and the word-meaning recall test, respectively. In fact, this condition created a large effect size of 1.88 on word-form acquisition and of 1.82 on word-form retention. These effect sizes in the case of word-meaning were even larger with a *d* value of 2.83 for both acquisition and retention. It is clear that these effect sizes are markedly larger than that of 0.41 in Rodgers' study (2013) and that of 1.38 in Vidal's study (2011). Study 6 shows that this condition also produced larger effect sizes on vocabulary gain than the NV condition in Study 4. There are several plausible explanations for this better lexical uptake. In the following part, I shall look at each explanation in detail.

First, this larger vocabulary gain can be attributed to the fact that the learners in the NVS condition obtained better text comprehension than those in the NV condition, as shown in Study 6. Following Krashen's Comprehensible Input Hypothesis (1985), the former were also hypothesized to make a larger vocabulary gain than the latter. This hypothesis is strongly supported by a positive and moderate correlation between the text comprehension levels and the size of vocabulary gain in these learning conditions. The Multiple Regression Analyses in Study 6 further confirm this hypothesis.

It should be also noted that the learners in the NV condition were merely required to view the videos with a view to answering the content questions afterwards. Meanwhile, those in the NVS condition were asked to give a classmate a summary of the video(s) that they had just viewed and listen to their classmate summarizing the content of the video(s) that they were going to view later on. Further analyses with the text comprehension scores obtained by the learners in these conditions show that asking the learners to give a summary of the video content after viewing and listen to a summary of the video content before viewing both led to significantly better text comprehension than merely having them view the videos. This finding supports Wittrock's Model of Generative Teaching of Comprehension (1991) regarding the positive effect of summarizing input content after listening on text comprehension. It is also consistent with what previous research (e.g., Herron, 1994; Herron

et al., 1998) has found about the benefit of using a summary as an advance organizer in the pre-listening stage to foster L2 listening comprehension.

However, it was receiving a summary of the video content before viewing rather than giving a summary of the video content after viewing that was found to help the learners with their incidental vocabulary acquisition. Receiving a summary of the video content may have served as an advance organizer of upcoming information, and this may have facilitated text comprehension. This, in turn, may have freed up more mental resources to deal with unfamiliar words in the input. This account is supported by a positive and moderate correlation between the quality of the summary that these learners received before viewing and the size of their vocabulary gain. On the other hand, when the students were processing video content with a view to summing it up to a fellow-student, they may have invested most of their mental resources for the purpose of overall text comprehension rather than attending to novel words. This is evidenced by the tendency for a negative correlation between the quality of their summary and the size of their own vocabulary gain. This finding is consistent with the widely accepted thesis that working memory has limited capacity (Cowan, 1995; Robinson, 2003; Schmidt, 1990).

As an output task, this summary activity also stimulated incidental vocabulary acquisition according to two major functions of Swain's Output Hypothesis (2005): fluency development function and hypothesis testing function. First, this activity provided the learners with opportunities to use the target words in their summaries and therefore helped them retain these lexical items until even two weeks after the experiment. Second, the contexts in which the target words were used were all different from the original contexts in the videos. Additionally, one learner was even found to use a target word with a new grammatical function that differed from its use in the input materials. These findings imply that the learners must have formed and tried out some hypotheses about how the target words should work in the new contexts and with the new grammatical function. However, it should also be acknowledged that the number of the learners who used any target words in their summaries (five out of the 30 learners) and the number of the target words that were actually used (two out of the 29 target words) were small. The incorporation of target words in the output activity should therefore not be overestimated as an explanatory factor for the learners' vocabulary gain in this experiment.

During the output task, the learners were free to interact with their interlocutors, and thus ask for clarification of the target words that their interlocutors used in their summaries. Learning through ‘negotiation-of-meaning’ is a key aspect of Long’s Interactional Hypothesis (1996). However, there was only one of such negotiation episode in the present study. Moreover, in this negotiation episode the summary giver had not acquired the precise meaning of the target words she was trying to use, and so she failed to clarify them adequately to her classmate. Neither student recalled these words in the post-tests.

The limited effect of the interactional output task on incidental vocabulary acquisition in the present study may be, again, associated with the nature of the task. It should be noted that previous research in this area used an output task which provided *focus on formS* (e.g., Ellis & He, 1999) or *planned focus on form* (de la Fuente, 2002). In fact, L2 learners in these studies were explicitly given the form and the meaning of the target words before they were required to use these words in the output tasks. However, the present study used an output task which focused on meaning and did not stipulate that students should use particular words. It was therefore the learners who decided whether to use the target words in their summaries, depending on how relevant (or ‘task-essential’) they felt them to be for their summaries. Moreover, successful resolution of negotiation-of-meaning episodes will depend on how well the learner has understood the given word. In fact, the learners in the present study may not have felt confident enough about using the newly encountered words, and may have avoided them in their summaries. In addition, when interaction was integrated into an output task in previous research (e.g., de la Fuente, 2002), L2 learners were often given the opportunity to interact with an interlocutor who had a noticeably higher L2 proficiency level, such as their teacher or a native speaker of the target language. In this way, the learners received the precise form or meaning of the target words from this interlocutor. By contrast, the learners in the present study interacted with their classmates, who did not always know the precise form or meaning of the target words. For all these reasons, it is perhaps not surprising that the interactional aspect of the output activity in this experiment was not found to benefit vocabulary learning much.

In the NVS condition, the size of vocabulary gain was also found to be positively and moderately correlated with the frequency of word occurrences in the input materials, the usefulness of verbal clues in these materials for word-meaning inference, and the prior L2 proficiency of the learners. This finding is consistent with those of Sawada (2009) and Vidal (2003, 2011).

In brief, the learners were indeed able to make a statistically significant vocabulary gain from this learning condition. This vocabulary gain was even larger than that in the NV condition. This better vocabulary gain was due to the fact that the learners in the NVS condition were given a summary of the video content before they actually viewed the video(s) themselves. As an advance organizer of upcoming information, this summary fostered their listening comprehension and incidental vocabulary acquisition at the same time. However, there was little evidence that the learners in the NVS condition enhanced the size of their vocabulary gain through using and negotiating the form and/or meaning of the target words in this interactional summary activity.

Study 7: A Qualitative View into the Benefits of the Output Tasks for Vocabulary Gain

This study aimed to investigate the cognitive and metacognitive processes that the learners used to deal with the output tasks in the RVS condition and in the NVS condition and how these processes benefited their text comprehension and incidental vocabulary acquisition. This investigation was carried out through my semi-structured one-to-one interviews with four learners who had participated in the RVS condition and another four learners who had undertaken the NVS condition. In what follows, I shall discuss the benefits of the summary tasks that the learners in these two conditions had reported and relate these benefits with dominant theories in the fields of L2 listening comprehension and incidental vocabulary acquisition: Wittrock's Model of Generative Teaching of Comprehension (1991), Long's Interaction Hypothesis (1996), Ausubel's Advance Organizer (1978), Hulstijn and Laufer's Involvement Load Hypothesis (2001), Nation's Vocabulary Generation (2013), and Swain's Output Hypothesis (2005). In general, the findings from this study seem to support all theories above. However, there are also several other findings that go beyond the core tenets of those theories.

As already mentioned in the Literature Review Chapter, Wittrock's Model of Generative Teaching of Comprehension (1991) included two main arguments that are directly relevant to the practice of using a summary activity as a generative tool to foster text comprehension. First, when learners are required to summarize input content, they need to generate mental connections between different ideas within the input content and between this content and their previous knowledge and experience. Second, this generation, in turn, helps them to improve their comprehension and memorization of this content.

In preparation for the summary tasks, the learners in the present study were found to skim their notes again, distinguish the main ideas in the listening input from the supporting ideas and factual information, and then incorporate only these main ideas into their summaries. This therefore suggests that these learners had learnt about the discourse features of a typical summary and now they used this prior knowledge to select only relevant information to include in their summaries. Put differently, through these activities, the learners did mentally associate the listening content with their relevant schemata. Another cognitive activity they carried out in this preparation stage was their organization of the main ideas in a coherent summary. To do so, they must have formed mental links between and among these ideas. Therefore, these findings together appear to support the first argument in Wittrock's Model of Generative Teaching of Comprehension (1991) mentioned above.

Apart from these cognitive processes, these summary activities also created a favourable condition for the learners in this study to exercise an orchestra of many metacognitive processes. For example, these activities gave them the opportunity to self-evaluate the quality of their previous listening and identify shortcomings in their current text comprehension. Interestingly, the learners often used these shortcomings as a basis to set goals or foci for their second listening. They also deployed the key words or key events in their summaries as signposts to facilitate their second listening. In other words, through these summary tasks, the learners reflected on their listening process and outcome, recognized their listening problems, and then sought for suitable strategies to deal with these problems in their second listening. According to the learners in this study, it was the combined effect of the cognitive and metacognitive processes mentioned above that helped them to enhance their comprehension and memorization of the listening content. Thus, these findings are also consistent with the second argument that Wittrock (1991) proposed in the model above.

The findings from the present study also echo what Thompson and Rubin (1996) and Vandergrift and Tafaghodtari (2010) have found in their experiments about the importance of metacognition in L2 listening comprehension. However, unlike in the case of the two studies above, the learners in the present study had not been trained about how to use metacognitive strategies to facilitate their L2 listening comprehension. Neither had they been advised to do so during the current experiment. It was the summary tasks themselves that stimulated the learners to implement these metacognitive processes.

This study in fact makes several other contributions to my understanding of Wittrock's Model of Generative Teaching of Comprehension (1991). First, it provides empirical evidence that this model can also be applied to L2 listening instruction. It should be noted in this regard that the present study is the first one to date that has examined the effect of using a summary activity as a generative tool on L2 listening comprehension. Second, in his latest model, Wittrock (1991) regarded metacognition as an external tool that can be used to complement generative learning. However, the findings from this study suggest that metacognition can be an integral component of generative learning. As can be seen from the discussion above, when the summary activities were incorporated into a meaning-focused listening lesson as in the case of the RVS or the NVS condition in the present study, these activities automatically triggered both cognitive and metacognitive processes from the learners even though these learners had not been trained or advised to do so when they participated in the output tasks. Finally, it was also confirmed by the learners in the present study that the summary tasks in the RVS condition and the NVS condition both facilitated their text comprehension, especially in the second viewing. With this better text comprehension, they were more able to make use of available contextual clues in the input materials for their lexical form-meaning mappings.

In his latest version of Interaction Hypothesis, Long (1996) posited that interaction gives learners opportunities to negotiate both the form (i.e., the language packaging) and the meaning (i.e., the content) of a given message, which, in turn, fosters language acquisition, including incidental vocabulary uptake. In the present study, the learners in the NVS condition engaged in four negotiation episodes about the content of the given summaries. According to these learners, these negotiation episodes fostered their text comprehension in two different ways. On the part of the summary receivers, these episodes gave them the opportunity to seek content clarification, and then set goals or foci for their actual listening. On the part of the summary givers, when they received a request for content clarification, they needed to reprocess their perceived information, and then restructure it in order to get their intended message across. However, these learners also reported that although these negotiation episodes made the content of the summaries more understandable, they did not pick up many words from these summaries. Instead, it was receiving a summary of the input content before listening that fostered their actual listening comprehension, which, in turn, created a favourable condition for their incidental vocabulary uptake. In this summary activity, I also found three instances in which the learners directly negotiated the form and/or

meaning of unfamiliar words (*burglar alarm*, *joystick*, and *lure*). In these instances, as the learners received the precise word forms and/or meanings from their interlocutors, incidental vocabulary acquisition did occur. However, the number of such instances in the present study was very small, and so altogether the interactional component of the activity made only a very small contribution to the overall vocabulary gain attested.

As mentioned above, the interactional summary activity in the NVS condition allowed the learners to receive a summary of the video content before viewing. According to these learners, they used the key concepts, the key events, and also the sequence of these key events from the summary above as an advance organizer to facilitate their actual listening comprehension. This finding therefore seems to be in line with what previous research (e.g., Ausubel, 1978; Herron, 1994; Herron et al., 1999; Jafari & Hashim, 2012) has found about the benefit of using a summary as an advance organizer to foster L2 listening comprehension.

As output tasks, the summary activities in the present study also stimulated incidental vocabulary acquisition in accordance with Hulstijn and Laufer's Involvement Load Hypothesis (2001), Nation's Vocabulary Generation (2013) and Swain's Output Hypothesis (2005). First, these tasks helped the learners to notice gaps in their lexical resources and then fill these gaps either by consulting their interlocutors in the NVS condition or by directing their attention to relevant lexis in the second viewing in the RVS condition. In other words, the learners indeed needed some words either to summarize the video content or understand a summary of the video content, and then attempted to search for these words from available resources (e.g., from their interlocutors or the given input materials in their next viewing stage). Evaluation was also evidenced in these tasks as can be seen in the following example: *"In the summary, I used "a machine" rather than "optical lure". It is because I just caught its sound in my first listening, but not yet its meaning. So I did not use it. In the second time, I tried to guess its meaning again. It is a device to test how sea animals react to its light, artificial light"* (RVS603). Together, these tasks did create a favourable condition for the learners to process and then learn new lexical knowledge through three main task components – Need, Search, and Evaluation – as Hulstijn and Laufer (2001) proposed in their Involvement Load Hypothesis.

In addition, through these output tasks, the learners did use words that they had encountered for the first time in their first viewing to summarize the video content. It is worth noting that these words were all used in syntagmatic contexts which differed from the

original phrasings in the given videos. As reported by the learners, such use helped them to create a durable memory of these lexical items. Their post-test scores gave more evidence for this. According to Swain (2005), in order for learners to use newly met words in novel contexts, they must have formed some hypotheses about how these words should work in those contexts and then tried out these hypotheses through meaning-focused output tasks. The learners in the present study also recounted that the opportunity to be exposed to the input material again in the RVS condition or interact with their classmates in the NVS condition, in some cases, indeed helped them to verify such hypotheses and then fine-tune their lexical knowledge. Together, these output tasks indeed brought about vocabulary gain according to four major functions that Nation (2013) and Swain (2005) consider the benefits of output tasks: noticing, fluency development, hypothesis testing, and generative use.

However, the findings from this study also make other contributions to my understanding of Swain's Output Hypothesis (2005). First, previous research in this area often used an output task which intentionally focused learners' attention on target words, i.e., a sequence where a meaning-focused activity (e.g., reading comprehension) was followed by a form-focused output activity (e.g., using selected target words from the reading input). However, in the present study, the whole activity sequence (including summarizing the listening content) was primarily meaning focused. In that sense, the vocabulary acquisition examined in this thesis is situated more clearly within the realm of incidental acquisition – i.e., acquisition as a by-product of meaning-focused activities. If the learners wished to use a target word in their summaries, they needed to have established a form-meaning mapping for this lexical item during their listening: *"I noted down here the word "plange" (an ill form of "plunge") as it seems to be important to understand the final part of the talk. However, I have not learnt this word yet so I did not use it and also removed the idea related to it from my summary"* (RVS601). In case of doubt, they would either leave it out or resort to paraphrase: *"In the summary, I used "a machine" rather than "optical lure". It is because I just caught its sound in my first listening, but not yet its meaning. So I did not use it. In the second time, I tried to guess its meaning again. It is a device to test how sea animals react to its light, artificial light"* (RVS603). It should be acknowledged that the output task in this learning condition did not have as strong an effect on incidental vocabulary acquisition as those tasks used in the previous studies which provided either *focus on form* (e.g., Ellis & He, 1999) or *planned focus on form* (e.g., de la Fuente, 2002). However, it was still conducive to this learning process through enhancing the learners' text comprehension and then fostering their

lexical form-meaning mappings, especially in the second viewing. Thus, this finding suggests that, apart from the output tasks that offer *focus on form* or *planned focus on form*, those that *focus on meaning* do provide *incidental focus on form* and then lead to incidental vocabulary uptake (albeit with a smaller effect size).

Second, in her Output Hypothesis, Swain (2005) suggested that when interaction is integrated into a text-based output task, L2 learners have the opportunity to talk about their knowledge of newly met words with their interlocutor, which is often referred to as the metalinguistic function of output tasks. In my monologue summary task, no learner-learner interaction was available. However, through this activity, the learners did reflect on their knowledge of newly met words, and then decided whether to use these lexical items in their summaries (see the interview excerpt about “*plange*” above). As already discussed above, this self-reflection prompted the learners to direct their attention to relevant lexis in their second listening and incidental vocabulary acquisition indeed occurred on this premise. In other words, this summary activity also stimulated the learners to carry out different metacognitive processes to deal with unfamiliar words in the input material. This function may be different from the metalinguistic function that Swain proposes in her Output Hypothesis (2005). In addition, this finding also suggests that interactional feedback is not the only way to help learners reflect and fine-tune their knowledge of newly met words. In fact, giving learners another exposure to the same input materials also fulfils this role.

Summing up, the findings from the present study are consistent with the core tenets in five major theories in the fields of L2 listening comprehension and incidental vocabulary acquisition: Ausubel’s Advance Organizer (1978), Hulstijn and Laufer’s Involvement Load Hypothesis (2001), Nation’s Vocabulary Generation (2013), Swain’s Output Hypothesis (2005), and Wittrock’s Model of Generative Teaching of Comprehension (1991). However, too few language-related episodes occurred in the NVS condition to attribute much vocabulary acquisition in this study to the workings of Long’s Interaction Hypothesis (1996). In addition, there were some findings different from what previous research has found for the theories above and therefore expanded my understanding of these theories. First, although the output tasks in the present study were entirely meaning-focused, they still provided *incidental focus on form* and thus led to incidental vocabulary acquisition. Second, these output tasks stimulated incidental vocabulary acquisition according to Nation’s Vocabulary Generation (2013), and Swain’s Output Hypothesis (2005), but the effect sizes of these tasks on lexical

uptake were likely to be dependent on how successfully the learners had already mapped the form to the meaning of unfamiliar words during their preceding listening.

Summary

The outcome of the discussion above can be summarized by the following points. First, L2 learners were able to pick up new vocabulary knowledge from a single viewing of one TED Talks video, which therefore supports the claim by Coxhead and Walls (2012), Floyd and Jeschull (2012), and Takaesu (2013) about the potential benefits of TED Talks for incidental L2 vocabulary acquisition. However, the size of this lexical uptake is relatively limited, at least, regarding the number of novel words being acquired. This reflects the slow nature of incidental vocabulary acquisition (Nagy, Herman, and Anderson, 1985) and also confirms the challenges of this learning process from listening input (van Zeeland, 2013), including TED Talks (Coxhead & Walls, 2012). Therefore, pedagogical interventions are required to enhance the size of vocabulary gain from this source.

All the interventions I put to the test in this research project were indeed found to help to enlarge the size of vocabulary gain. There are two crucial factors that accounted for this better lexical uptake. As expected, the first and also the more important factor is text comprehension. In fact, having L2 learners view the same TED Talks video twice led to better text comprehension than having them view this video only once. This is consistent with what previous research has found about the positive effect of repeated listening on text comprehension (e.g., Chang & Read, 2006; Saika, 2009). Narrow Viewing indeed fostered text comprehension, which thus supports Krashen's Narrow Listening (1996). The summary activities in this research project also helped the learners in the RVS condition and the NVS condition to obtain better text comprehension than those in the RV condition and the NV condition, respectively. This finding therefore supports Wittrock's suggestion of using retelling as a generative tool to foster text comprehension (1991). However, the findings from this research project also indicate that when this tool was used in such a way as in the case of the RVS condition and the NVS condition, it prompted the learners not only to generate mental connections between different ideas in the input content and between these ideas and their relevant schemata but also to carry out metacognitive processes to facilitate their following listening. Thus, metacognition might be an integral component of generative learning rather than merely a facilitative external factor as Wittrock (1991) claimed. With this better text comprehension, the learners were more able to make use of available contextual

clues in the input materials for their lexical form-meaning mappings. This is in line with Krashen's Comprehensible Input Hypothesis (1985).

The other important factor is how the learners dealt with the target words in their output tasks. In this research project, the output tasks were also found to stimulate incidental vocabulary acquisition in accordance with Hulstijn and Laufer's Involvement Load Hypothesis (2001), Nation's Vocabulary Generation (2013), and Swain's Output Hypothesis (2005). However, the amount of lexical knowledge gained in this way heavily depended on the success rate at which the learners had inferred the meaning of the target words during their preceding listening and whether these words were relevant to their summaries. In general, the vocabulary gain that the learners in this research project made through using and negotiating the target words in these output tasks was relatively limited. This may be attributed to the nature of these output tasks, which, unlike the text-based output activities used in some previous studies, did not explicitly focus the learners' attention on pre-selected words by stipulating these should be used in the output activity. Instead, the entire activity sequence used in the present experiments stayed meaning-focused (instead of adding a form-focused phase), and thus more clearly situated in the realm of incidental acquisition.

In this research project, the interventions that integrated a meaning-focused output task after a meaning-focused input task led to a better vocabulary gain than the interventions that included a meaning-focused input task alone. This better vocabulary gain was found not only in terms of word meaning but also in terms of word form. It should be noted here that the learners in this research project were all informed beforehand about their assigned experimental procedures (but not the post-tests). Thus, when they knew that they were going to complete a text-based output task after their first listening, this is likely to have influenced the way they processed the given input. To be more specific, they may have processed both the form and the meaning of unfamiliar words which they might need to use in their output task. Since the output tasks in this research project required the learners to summarize the input content, they seemed to fasten their focus on the key words rather than on the other lexical items in the input materials. This suggests that anticipating a particular follow-up text-based output task may indeed influence the way learners process the given input text.

Finally, this research project has consistently provided supporting evidence for the crucial role of the frequency of word encounters and usefulness of verbal clues in the input

materials for word-meaning inference in incidental vocabulary acquisition from listening input.

Implications

Implications for Instructional Practice

Pedagogical values of the interventions

The five interventions that I suggested and put to the test in this research project were all found to bring about statistically significant vocabulary gains. Thus, they can be integrated into L2 listening courses to foster incidental vocabulary acquisition. However, if we view these gains from the numbers of novel words that the learners in this research project indeed picked up from their viewing procedures, these gains turn out to be relatively modest. From a pedagogical perspective, some may therefore question the value of these interventions. Still, it can be argued that these classroom procedures actually have many pedagogical merits.

First, it needs to be reiterated here that incidental L2 vocabulary acquisition is characterized by small increments (e.g., Herman et al., 1987; Nation, 2013; van Zeeland & Schmitt, 2013; Webb, 2008). Put differently, when learners encounter a novel word in a meaning-focused input activity for the first time, they may grasp only some portion of its form, meaning and/or use. Their knowledge of this lexical item will be further developed when they have opportunities to meet this word again in other contexts or to experiment with using this word in communication, and then receive feedback from their interlocutors on this use (Nation & Webb, 2011). Therefore, it seems irrational to expect incidental uptake of many words after a single lesson. However, if my proposed interventions were applied into five 45-minute lessons per week for 30 weeks per academic year as in the case of the language program where I collected the data for this research project, learners could theoretically pick up about 300 word families (two word families per lesson) from these interventions alone.

In addition, the gain attested in the vocabulary tests I used in this research project is likely to underestimate the amount of acquisition that actually took place. For one thing, the learners in the research project may have acquired partial knowledge of the target words, but this was not captured by the scoring protocol. For another, they may have acquired, fine-tuned, or simply consolidated knowledge of other words than those selected for the tests (as in the case of the phrase “*burglar alarm*” in Study 7). The fact that my vocabulary test was

not designed to gauge acquisition beyond the selected target words is another limitation of this study. Finally, the removal of the target words that were already known to more than 10% of the learners prior to the experiments, which I did to ensure an equal starting point between the treatment group and the control group regarding their entering knowledge of these lexical items, reduced the size of the data sets and thus probably the size of actual vocabulary gain.

The value of my proposed interventions not only lies in the quantity of novel words that the learners learnt but also the quality of their acquired lexical knowledge. Most previous research in this field has reported a quick attrition of L2 vocabulary knowledge incidentally acquired from meaning-focused listening activities (e.g., Brown et al., 2008; Vidal, 2003, 2011). However, the vocabulary gain in my interventions proved to be resistant to decay until, at least, two weeks after the treatment. As discussed in Section 1 of this chapter, there were several factors in these interventions, especially the presence of the output tasks in the case of the RVS and the NVS condition, which may have helped the learners to create a durable memory for their attained vocabulary knowledge. Admittedly, the fact that the size of vocabulary gain remained almost the same from the immediate post-test to the delayed post-test might be also attributed to the floor effect of such small gains in my interventions. However, the learners themselves confirmed the positive effects of such output tasks on their vocabulary retention in the follow-up interviews in Study 7. In addition, the effectiveness in which the learners used these words in their summaries regarding the grammatical accuracy, semantic and pragmatic appropriateness, and creativity provided a clearer evidence for the pedagogical value of my interventions for incidental vocabulary acquisition and retention.

Apparently, pedagogical interventions that are integrated in meaning-focused L2 listening lessons are mainly to facilitate learners in their processing of input content. My proposed interventions indeed effectively fulfilled this role. Apart from helping the learners to improve their text comprehension, these interventions also brought about noticeable vocabulary gains as an added value to the lessons. In other words, we, as L2 teachers and course designers, can achieve two instructional goals in the same lesson with no or little additional investment of classroom time. In these classroom procedures, learners also have opportunities to further develop their speaking ability, note-taking and summarizing skills, and other metacognitive strategies.

It should be also stressed that incidental vocabulary acquisition is merely one among many other ways of vocabulary learning. Nation (2001, 2013) convincingly put that incidental vocabulary acquisition and intentional vocabulary learning should be viewed as complementary activities: one enhances the learning that comes from the other. He further added that a well-designed L2 course needs to balance opportunities for learners to learn from both meaning-focused and form-focused activities. With the proposed interventions in this research project, I would like to suggest one useful way to optimize incidental L2 vocabulary uptake from meaning-focused listening activities rather than nominate these interventions as a substitution for the other ways of vocabulary learning, including such effective intentional vocabulary learning methods as flashcards or mnemonic techniques (Nation & Webb, 2011). Although I carried out all five interventions in the language classroom context in this research project, these interventions can also be transferred to other learning contexts like independent study in language learning centres, clubs or at home.

So far, I have pointed out some main pedagogical merits of my proposed interventions for incidental L2 vocabulary acquisition. To help L2 teachers and course designers to make full use of these interventions, I shall give some further suggestions regarding how to select input materials and carry out these interventions in real practice in the following sections.

Selecting suitable TED Talks

The findings from this research project show that TED Talks are a source of incidental L2 vocabulary acquisition. However, the size of vocabulary gain from this source proves to be relatively limited, especially when no pedagogical interventions are applied. To optimize the benefits of these materials for incidental vocabulary uptake, L2 teachers and course designers may consider the following factors in their selection of input materials and learners who will be exposed to these materials.

First, this research project has provided a consistent finding that the size of vocabulary gain from TED Talks was positively associated with the level of text comprehension. Previous research shows that text comprehension is, in turn, dependent on learners' prior L2 proficiency and vocabulary knowledge (e.g., Stæhr, 2008, 2009; van Zeeland & Schmitt, 2012). Together, these logically entail that the higher L2 proficiency and the larger vocabulary knowledge learners have before they view TED Talks, the better text comprehension and therefore the larger vocabulary gain they are likely to obtain from these input materials. The learners in my research project had an upper-intermediate level of

English language proficiency and a receptive knowledge of at least 4,000 most frequent word families in English prior to the experiment. However, it has been clear from this research project that they nevertheless struggled to take in the input content and pick up the form and meaning of novel words from the input materials. Thus, it might be a better idea to use TED Talks with learners who have an advanced level of English language proficiency and a vocabulary size of between 8,000 and 9,000 most frequent word families in English. In fact, according to Coxhead and Wells (2012), with a vocabulary size from 8,000 to 9,000 most frequent word families in English, learners will benefit from a 98% lexical coverage of most TED Talks – a threshold that is generally believed sufficient enough to help them cultivate an adequate understanding of L2 listening input. When learners have such L2 proficiency and vocabulary knowledge, they are expected to better understand the input content and therefore find it easier to guess (and retain) the meaning of the few remaining words that they are not yet familiar with.

The results from the Narrow Viewing condition suggest that the more the learners were familiar with the topic of the input text, the better they tended to understand the text and the better they tended to acquire new words. Thus, if a L2 listening lesson aims not only to foster text comprehension but also to facilitate incidental vocabulary uptake from this listening procedure, L2 teachers and course designers may need to select a listening topic that is familiar to their learners and ideally relevant to their learners' own interest. Research in the context of L2 reading shows that familiar and interesting topics benefit both text comprehension and incidental vocabulary acquisition (e.g., Lee & Pulido, 2017; Pulido, 2003). This might also hold true in the case of L2 listening. On the website of TED Conference, there is a list of topics which is organized in an alphabetical order. Learners may be asked to rate the familiarity and relevance of these topics to their own interests at the beginning of their L2 listening course. From their ratings, the most familiar and favourite topics could then be incorporated into the course.

As already mentioned in the Literature Review Chapter, Scotto di Carlo (2014) found from his analysis of 80 TED Talks of various disciplines that the presenters in these talks often use metaphors, comparisons, and/or exemplifications to explain the meaning of technical words. In the case of the present research project, the results show that when a novel word was saliently surrounded by verbal clues for word-meaning inference, this lexical item stood a better chance to be learnt. Therefore, if L2 teachers and course designers would like to create a favourable condition for their learners' incidental vocabulary uptake from

TED Talks, they may select the talks in which their target words are surrounded by useful verbal clues for word-meaning interpretation. As each TED Talks video is accompanied by an English transcript, they can first copy and paste this transcript into a Microsoft Word file. Subsequently, they may use the FIND function in the Microsoft Word program to locate the target words and rate the usefulness of verbal clues in the surrounding context for word-meaning inference.

Another factor that was also found to facilitate incidental vocabulary acquisition in this research project is the frequency of word occurrences in TED Talks: the more often a novel word occurs in the input materials, the more likely it was to be picked up. Therefore, TED Talks in which the target words occur more frequently (and ideally in various contexts) should be chosen. The FIND function in a Microsoft Word program is also able to indicate the frequency at which these target words occur in each transcript. Alternatively, Nation and Heatley's Range Program (2002) also enables L2 teachers and course designers to compare the frequency of word occurrences across different transcripts. This program is available in Professor Paul Nation's webpage⁴ or in Tom Cobb's Lextutor website⁵.

Summing up, it is obvious from this research project that L2 learners were indeed able to pick up new lexical knowledge from viewing TED Talks. Thus, L2 teachers and course designers might incorporate these materials in their L2 listening courses, especially for advanced learners who have receptive knowledge of 8,000 to 9,000 most frequent word families in English. However, to optimize the benefits of these materials for their learners' lexical uptake, L2 teachers and course designers are advised to select the talks that have familiar and interesting topics to these learners and that have the target words reoccurring frequently and/or surrounded by useful verbal clues for word-meaning inference.

Designing classroom procedures

Repeated viewing

The results from Study 2 show that having L2 learners view a TED Talks video twice fostered better text comprehension and incidental vocabulary gain than having them view the same video only once. Thus, a repeated viewing of the same TED Talks video is the first classroom procedure that I advise L2 teachers and course designers to include in their L2

⁴ <http://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-programs>

⁵ <http://www.lextutor.ca/range/>

listening courses. There are two factors that might help L2 learners to make full use of this classroom procedure for their incidental vocabulary acquisition. First, learners might need to be informed beforehand that they will have more than one time of viewing the video(s). This announcement may reduce their listening anxiety and therefore get them more cognitively engaged with their listening process. Second, the results from Study 7 show that some learners noted down some gaps in their lexical resources onto the note sheets, and then attempted to fill these gaps in their second viewing (see Appendix I for an example). Thus, note-taking should be used in this classroom procedure. In fact, this technique may be crucial for incidental vocabulary acquisition from listening input in general and from TED Talks in particular, given the limited capacity of human working memory, the real-time nature of listening, and the presence of many technical words in TED Talks.

However, it should also be acknowledged that having L2 learners listen to the same text several times is not an authentic task in a sense that learners may rarely encounter such a task in their real-life listening. In addition, this listening/viewing procedure might also lead to their loss of interest.

Repeated viewing plus output

The second classroom procedure that this research project has also found to be conducive to L2 listening comprehension and incidental vocabulary acquisition is to have L2 learners view a TED Talks video, make an oral summary of the content, and then view this video again. This classroom procedure indeed brought about even a larger vocabulary gain than having learners merely view this video twice.

As already mentioned in Section 1.3 and 1.6, the anticipation of a follow-up text-based output task prompted the learners in this research project to process and pick up both the form and the meaning rather than the meaning of the target words only during their first viewing stage. In addition, when the learners knew that they were required to summarize the input content after their first viewing, they tended to fasten their focus on the key words rather than the other lexical items in the input materials. Thus, if a L2 meaning-focused listening lesson aims to facilitate incidental uptake of both the form and meaning of key words in the listening input, teachers and course designers might need to inform their learners about the presence and the nature of their follow-up text-based output task before exposing them to the input materials for the first time.

The results from this research project have shown that the benefit of this output task for incidental uptake of key words was only maximized when L2 learners had already succeeded in their inference about the meaning of these lexical items during their preceding viewing. Therefore, TED Talks in which the key words are saliently surrounded by useful verbal clues for word-meaning interpretation and reoccur frequently in various contexts should be selected for this classroom procedure. In Study 3, a summary activity was used as a listening-based output task. However, previous research (e.g., Rassaei, 2015) shows that such other text-based output tasks as composing comprehension questions based on the input content also foster incidental vocabulary acquisition, especially when L2 learners are encouraged to use newly met words in their compositions. Thus, these output tasks could also be inserted between two times of viewing the same video.

Again, note-taking should be used in this classroom procedure for three main reasons. First, this technique helps learners to overcome the limited capacity of human working memory and the challenges of real-time listening. Second, it provides learners with information for their summaries. Finally, it also allows them to note down lexical gaps, and then prompts them to fill these gaps in their second viewing.

On the downside, this classroom procedure is not authentic in the regard that learners have the opportunity to view the same video again and they are required to make a monologue summary of the video content. However, given the benefits it brings about for both L2 listening comprehension and incidental vocabulary gain, this procedure does have pedagogical values.

Narrow viewing

The results from Study 4 show that having L2 learners view three TED Talks videos on similar topics was beneficial for both text comprehension and incidental vocabulary acquisition. Therefore, this classroom procedure also deserves a place in L2 listening courses. To optimize the positive effect of this classroom procedure on vocabulary gain, L2 teachers and course designers might consider following factors.

This research project has consistently indicated that the size of vocabulary gain was positively correlated with the degree of text comprehension. Previous research has shown that text comprehension is, in turn, dependent on L2 learners' lexical coverage. Thus, when L2 teachers and course designers would like to have their learners view several TED Talks

videos on similar topics, they are recommended to choose the talks that have as many unfamiliar words in common as possible. In this way, they can reduce the number of unfamiliar words that their learners will encounter when viewing these videos together in the same lesson (or sequence of lessons, or materials for independent study). Ideally, these unfamiliar words should reoccur frequently in various contexts in the selected TED Talks.

In addition, there are two possible ways to sequence these videos with an aim to facilitate incidental vocabulary acquisition. First, L2 teachers and designers may arrange these videos in an order from the easiest to the most difficult one in terms of their listening content. In this way, learners can use their understanding of an easy video to aid their comprehension of a more difficult one. When their listening comprehension is mentally catered for, they will have more attentional capacity to deal with unfamiliar words in the following videos. Alternatively, the video in which the meaning of unfamiliar words is the most easily inferred from the surrounding context should be placed first. When learners have succeeded in their interpretation of these words, this will help them to reduce the number of unfamiliar words in the following videos. In other words, they can increase their lexical coverage when they move on from one video to another.

Like in the case of the two classroom procedures discussed above, note-taking should also be used in the present classroom procedure. In addition, if the total viewing time is rather long, learners may need to have some break in between in order for them to make any modifications and/or additions to their notes and also to avoid their loss of concentration.

Narrow viewing plus output

The final classroom procedure that I recommend L2 teachers and course designers to include in their listening courses is to have their learners view either the first or the second half of a set of TED Talks videos on similar topics, share the content to a classmate who has not viewed the same half yet in an interactional summary activity, and then move on with the other half. The results from Study 5 show that this classroom procedure brought about even better text comprehension and vocabulary gain than having learners merely view this set of videos once.

The results from Study 5 and 7 suggest that if the target words were the key words in the input materials, these words were more likely to be learnt. It is because key words tend to trigger L2 learners' attention when they process a given input text for overall comprehension

and also when they summarize and negotiate the input content with their classmates. Thus, L2 teachers and course designers may need to select TED Talks in which the words they would like their learners to pick up are also the key concepts of the talks. Ideally, these words should be accompanied by useful contextual clues for word-meaning inference and/or reoccur frequently in those talks.

In the Narrow Viewing plus Summary condition in this research project, the size of vocabulary gain was found to be positively correlated the quality of the summary of the video content that the learners received from their classmates before they themselves viewed the other half of the video set. Therefore, it might be a good idea to pair a strong learner with a weak learner in this classroom procedure. The former should be required to view and summarize the half with a more difficult listening content while the latter does the same, but with an easier one. In this way, the weak learner can also enjoy the benefits that this classroom procedure brings about for their text comprehension and incidental vocabulary acquisition.

As discussed above, one possible explanation for the small effect of the output task in the Narrow Viewing plus Summary condition on vocabulary gain is the learners' avoidance of using the target words in this task. Therefore, in this suggested classroom procedure, L2 learners should be encouraged to use newly met words even when they are not completely sure about the knowledge of these words yet. This is because, through this use, they are able to try out their hypotheses about how these words should work in context and possibly confirm, modify or reject these hypotheses based on their classmate's feedback. In case their classmate's feedback cannot help them to verify those hypotheses, this word use may, at least, prompt their classmates to attend to these lexical items when they view the video(s) themselves.

The results from Study 3 and 7 show that when the learners were given another viewing of the same TED Talks video after their text-based output task, they were able to fill the lexical gaps that they recognized during the output task or compare their own output with the given input and then fine-tune their lexical knowledge. Thus, when L2 teachers and course designers would like to use the present classroom procedure (NVS) in their listening courses, it might be a good idea to ask their learners to watch the video they have just summarized again together with their classmates who view this video for the first time. This

classroom procedure has been suggested by Boers et al. (2007) and Lindstromberg and Boers (2008).

Again, L2 learners in this classroom procedure should also be informed beforehand about the presence and nature of the follow-up listening-based output task in order to prompt them to process both the form and the meaning of unfamiliar words. In addition, note-taking is also recommended in this classroom procedure to help learners to deal with the limited capacity of human working memory and the real-time nature of listening.

Training necessary learning strategies

In this research project, the learners were not trained about how to take notes of listening content, summarize this content, or use metacognitive strategies to foster their listening comprehension. The results from this research project, however, have indicated that these learning strategies benefited not only text comprehension but also incidental vocabulary acquisition. If these learners had been trained about these strategies, the positive effects that these strategies had on their incidental vocabulary uptake could have been larger.

Implications for future research

The findings from this research project suggest that, in order for L2 learners to learn an unfamiliar word through a combination of a meaning-focused input task and a meaning-focused output task, they first need to notice the presence of this word in the input materials (*noticing*). In addition, this word has to play a crucial role for their understanding of the input content (*motivation*). Otherwise, it will be ignored. Subsequently, learners need to retain this word in their working memory longer than usual in order to implement the process of form-meaning mapping for this item (*retention*). When they have opportunities to retrieve, use and negotiate this word (with their interlocutors) in the follow-up meaning-focused output task, they can not only refine their knowledge of this word but also create a durable memory for this item (*further noticing, retrieval, and generation*). The aforementioned learning conditions of *noticing, motivation, retrieval, generation, and retention* are exactly the same as what Nation and Webb used to frame their Technique Feature Analysis (2011). Like Hulstijn and Laufer's framework of Involvement Load Hypothesis (2001), this framework was also constructed to estimate the cognitive involvement load that a vocabulary learning task induces from L2 learners and therefore the relative effect that this task may produce on vocabulary gain. Hu and Nassaji (2016) found that this framework had a significantly better

predictive power for the size of vocabulary gain from a learning task than the Involvement Load Hypothesis. Thus, future research may use Nation and Webb's Technique Feature Analysis framework to guide their construction of vocabulary learning tasks in their experiments. It might be also interesting to compare the effects on the size of vocabulary gain of different tasks which have different degrees of noticing, motivation, retrieval, generation and retention and retention as the way Kim (2008) and Hulstijn and Laufer (2001) compared the effects of tasks with different degrees of need, search and evaluation in their validations of the Involvement Load Hypothesis framework.

According to Wittrock's Model of Generative Teaching of Comprehension (1991), using a summary activity as a text-based output task prompts learners to generate mental connections between different ideas in input content and between these ideas and their schemata and this generation, in turn, fosters their comprehension and memorization of this content. In addition, metacognition is an external factor that can assist this generative learning. The results from this research project support this model. However, when such a summary activity was integrated between two times of viewing the same video as in this research project, this activity naturally prompted the learners to carry out different metacognitive strategies to facilitate their following viewing, which therefore enhanced their text comprehension. In other words, metacognition might be an integral component rather than merely an external facilitative factor of generative learning. Nevertheless, as this research project is the first one that has investigated the effect of a summary activity as a generative tool to foster L2 listening comprehension, we still need more empirical evidence to support the above account. Therefore, I call for more empirical studies in this area.

The results from this research project have also provided supporting evidence for Nation's Vocabulary Generation (2013) and Swain's Output Hypothesis (2005). However, this research project indicates that not only the output tasks that provide opportunities for *focus on form* or *planned focus on form* (as often found in previous research) but also those that mainly *focus on meaning* (the listening content) and offer opportunities for *incidental focus on form* are conducive to vocabulary gain (albeit with a smaller effect size). In addition, L2 learners are able to refine their lexical knowledge not only through interactional feedback but also through comparing their own output against the given input. Therefore, it might be interesting for future research to compare the effect on the size of vocabulary gain between these types of output tasks.

Regarding the research design, this research project is distinctive in this field in several ways. It is the first study to use follow-up interviews with learners to examine the effects of the listening-based output tasks on their vocabulary learning. The data from these interviews indeed have helped me to understand more about the cognitive and metacognitive processes that the learners used to deal with the output tasks and how these processes affected their text comprehension and incidental vocabulary acquisition.

The vocabulary measures in this research project were arguably better suited than those in most of previous research in the regard that they used the spoken form of the target words as the test prompts. This therefore helped to ensure the congruence in the way the learners were exposed to the target words in the input materials and in the vocabulary measures. This research project has developed an oral word-form recall test – a novel vocabulary measure in the field, even though the original idea for such a test was borrowed from Laufer and Nation's frequency-based productive written vocabulary knowledge test (1995). The quality of this test, together with the other dependent measures in this research project, was successfully validated using Messickian construct validity as a theoretical framework and Rasch model as an analytical tool. Therefore, this test might be an option to consider in future research in this field.

It should also be acknowledged that this research project has several limitations in its methodology. First, this research project used a between-participant research design to compare the degree of text comprehension across different treatment conditions. However, I only ensured that there was no difference in the learners' general L2 proficiency, but not their L2 listening ability between these conditions. My between-group comparison of text comprehension would have been more valid if I had administered a validated L2 listening ability test before these learners were assigned to treatment conditions. Scores from such a test could also have been used as a basis for examining the external validity of the text comprehension tests used in the present research project.

Second, I used Nation and Beglar's Vocabulary Size Test (2007) rather than Schmitt, Schmitt and Clapham's Vocabulary Levels Test (2001) to gauge the prior vocabulary size of the learners. Although a learner in this research project was only considered to have receptive knowledge of a 1,000 frequency-based vocabulary level when they had nine correct answers out of the total ten test items in each level, the nature of Nation and Beglar's Vocabulary Size Test does not in principle justify such a generalization. However, it should also be noted that

the learners' scores on this test were only used as a learner-related factor to compute the Pearson correlation coefficients with the size of their vocabulary gain rather than as a basis to select the target words in this research project. Therefore, this does not constitute a serious threat to the validity of my research methodology.

Regarding the input materials in this research project, while the learners in the pilot study all rated the three experimental videos as comprehensible and interesting the results from the main studies reveal that the contents of these videos were far from straight-forward to the research participants. As the learners had to invest most of their working memory capacity for the purpose of text comprehension, their remaining mental resources might have been inadequate for them to deal with unfamiliar words in the input materials.

Although the word-form recall test in this research project suggests that there is another validated way to measure the size of incidental L2 word-form uptake from meaning-focused listening activities, the way the learners were required to respond to the test items in this research project proved to be rather complicated. This, in turn, might have restricted their ability to demonstrate their true vocabulary knowledge.

The text comprehension measure in this research project is also limited in the regard that it included only one question type – Yes/No questions. It is undeniable that this question type provided the learners in this research project a fifty percent of correct answer merely by random guessing. Fortunately, the test validation in the Methodology Chapter shows that the number of such random guesses was too small to threaten the validity of the test as a whole.

As the qualitative investigation (i.e., Study 7) was carried out with another group of participants (in New Zealand), who had a noticeably higher level of English proficiency than those in the quantitative studies (in Viet Nam), no direct links could be made between its findings and those from the quantitative ones. While it seems plausible that the students who participated in the quantitative investigations engaged in processes similar to what was observed in the qualitative study, the fact that different learner profiles were involved calls for caution when it comes to using the qualitative data to shed light on the quantitative findings. This is another limitation of the present research project.

Last but not least, the sample size per treatment condition in this research project was relatively small, which restricts the generalizability of the research results to a larger population. In addition, previous research has suggested that other types of text-based output

tasks such as generating comprehension questions for the input content or predicting how this content can be further developed are conducive to vocabulary learning. However, I only had the learners in this research project use one kind of text-based output activity – an oral summary task. Therefore, I would like to call for more empirical research in this field to address the aforementioned limitations.

CONCLUSION

I started this research project with two core arguments. First, L2 teachers and course designers can add value to their meaning-focused listening lessons through creating opportunities for incidental vocabulary uptake. Second, one way to do so is supporting learners with their text comprehension. In this way, learners are expected to have more mental resources to process novel words in listening input. Using relevant theories and research findings in the fields of L2 listening comprehension and incidental vocabulary acquisition, I constructed five pedagogical interventions (One-time Viewing, Repeated Viewing, Repeated Viewing plus Summary, Narrow Viewing, and Narrow Viewing plus Summary) which were all deemed to foster text comprehension and therefore incidental vocabulary acquisition.

When these interventions were put to the test in the classroom experiments in this research project, the results showed that L2 learners were indeed able to make statistically significant vocabulary gains from these interventions. To be more specific, the learners in the One-time Viewing intervention made a mean gain of 0.5 target words. On average, the learners in the Repeated Viewing intervention and the Narrow Viewing intervention acquired one target word while those in the Repeated Viewing plus Summary intervention and the Narrow Viewing plus Summary intervention picked up two. As expected, text comprehension was found to significantly predict the size of vocabulary gains within and between the interventions. As compared to the One-time Viewing intervention, the Repeated Viewing intervention and the Narrow Viewing condition brought about better text comprehension thanks to the benefit of viewing the video a second time and viewing the videos of similar topics, respectively. This better text comprehension, in turn, assisted the learners in their lexical form-meaning mappings. As compared to the Repeated Viewing intervention and the Narrow Viewing Intervention, the Repeated Viewing plus Summary intervention and the Narrow Viewing plus Summary intervention also led to better text comprehension thanks to the benefit of the summary activities, where the learners had opportunities to reprocess perceived information from the previous viewing and get prepared for the following viewing. Again, this better text comprehension put the learners in a better position to infer the meaning of unfamiliar words out of context. These findings appear to support Ausubel's Advance Organizer (1978), Krasken's Input Hypothesis (1985), and Wittrock's Model of Generative Teaching of Comprehension (1991).

As listening-based output tasks, the monologue summary activity in the Repeated Viewing plus Summary intervention and the interactional summary activity in the Narrow Viewing plus Summary intervention benefited incidental vocabulary acquisition in three different ways. First, these activities helped the learners to notice gaps in their lexical resources, and then prompted them to fill these gaps by directing their attention to relevant lexis in their later exposure to L2 input or seeking for help from their interlocutors. Second, the activities also gave them opportunities to use newly met words in new contexts, negotiate the form and meaning of these words with their interlocutors, and therefore consolidate their knowledge of these lexical items. Finally, through this use and negotiation, the learners also reflected on their vocabulary knowledge, identified the problems when this knowledge was put into use, and then sought for the solutions to these problems. In this way, they might have developed their metacognitive strategies in lexical processing in meaning-focused input and output tasks. Together, these findings seem to be consistent with Hulstijn and Laufer's Involvement Load Hypothesis (2001), Nation's Vocabulary Generation (2013), and Swain's Output Hypothesis (2005).

It is also worth noting here again that the limited vocabulary gain in this research project may be mainly due to the nature of the input materials. Unlike most of previous research in this field, the present research project used authentic TED Talks as the input texts. As already mentioned, such inherent features of authentic listening materials as natural and fast manner of speech delivery might have, to some extent, restricted the learners' incidental vocabulary uptake. If the interventions above had been used with simplified input materials, the size of vocabulary gain might have larger.

From this research project, I learnt many valuable lessons about the nature of incidental L2 vocabulary acquisition from meaning-focused listening activities. First, this learning process is challenging due to many unavoidable factors, such as: the real-time nature of listening or the limited capacity of human working memory. Second, as a L2 teacher and course designer, I can help my learners to overcome these challenges by developing meaning-focused listening activities that reduce the cognitive demands for their text comprehension. In this way, my learners can have more attentional capacity to process novel words in listening input. However, for incidental vocabulary acquisition to occur, my learners need to direct their attention to these words, and then attempt to map the word form to the word meaning. This mapping proves to be facilitated if the words reoccur several times in input materials and are surrounded by useful contexts for word-meaning inference. My learners may further

develop their knowledge of these lexical items if they have opportunities to use and negotiate the form and meaning of these items with their interlocutors in listening-based output tasks. Therefore, such factors as *noticing* (Robinson, 2003; Schmidt, 2001), *motivation* (Hulstijn & Laufer, 2001), *comprehensible input* (Krashen, 1985; Long, 1985), *retrieval and retention* (Nation & Webb, 2011), *generation* (Nation, 2013; Swain, 2005), and *negotiation* (Long, 1996) are the key ingredients for my recipe of incidental vocabulary acquisition from meaning-focused listening activities. Nevertheless, it should be also noted that the specific recipe or, in other words, the proportion in which the ingredients above are combined can vary from lesson to lesson, depending on how creative and effective I am as a cook.

REFERENCES

- Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 51(5), 267.
- Ausubel, D. P. (1978). In defence of advance organizers: A reply to the critics. *Review of Educational Research*, 48(2), 251-257.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Baghaei, P., & Amrahi, N. (2011). Validation of a multiple choice English vocabulary test with the Rasch model. *Journal of Language Teaching and Research*, 2, 1052-1060.
- Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, 50, 507-31.
- Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, 50, 507-31.
- Beglar, D. (2009). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 26(4), 1-22.
- Berne, J. (1995). How do varying pre-listening activities affect second language listening comprehension?. *Hispania*, 78(2), 316-329.
- Boers, F., Eyckmans, J., & Godfroid A. (2007). Catering for limited processing capacity to foster incidental vocabulary uptake. In K. Pelsmaekers & C. Rollo (Eds.), *Economically speaking: Essays in honour of Chris Braecke* (pp. 169-187). Antwerp: Garant Publishers.
- Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14(1), 14-31.
- Brown, G. (1995). Dimensions of difficulty in listening comprehension. In D. Mendelshon & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 59-73). San Diego: Dominie Press.
- Brown, R. A. (2007). Enhancing motivation through repeated listening. *Information & Communication Studies*, 37, 9-15.

- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20(2), 136-163.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Caliendo, G., & Compagnone, A. (2014). Expressing epistemic stance in university lectures and TED Talks: A contrastive corpus-based analysis. *Lingue e Linguaggi*, 11, 105-122.
- Cervantes, R., & Gainer, G. (1992). The effects of syntactic simplification and repetition on listening comprehension. *TESOL Quarterly*, 26(4), 767-770.
- Chang, A., & Read J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly*, 40(2), 375-397.
- Chapelle, C. (1988). Construct definition and validity inquiry in SLA research. In L.F. Bachman & A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). New York: Cambridge University Press.
- Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31(5), 693-713.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.
- Coxhead, A. J., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *The TESOLANDZ Journal*, 20, 55-58.
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary. *Studies in Second Language Acquisition*, 24(1), 81-112.
- di Carlo, G. S. (2014). The role of proximity in online popularizations: The case of TED Talks. *Discourse Studies*, 16, 591-606.
- Doctorow, M., Wittrock M. C., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology*, 70(2), 109-118.
- Danan, M. (1992). Reversed subtitling and dual coding theory: new directions for foreign language instruction. *Language Learning*, 42(4), 497-527.
- Dupuy, B. C. (1999). Narrow listening: An alternative way to develop and enhance listening comprehension in students of French as a foreign language. *System*, 27(3), 351-361.

- d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28(3), 227-244.
- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, 89(2), 206-220.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.
- Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam: John Benjamin Publishing Co.
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21(2), 285-301.
- Ellis, R., & Heimbach, R. (1997). Bugs and birds: Children's acquisition of second language vocabulary through interaction. *System*, 25(2), 247-259.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, 44(3), 449-491.
- Etemadi, A. (2012). Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition. *International Journal of English Linguistics*, 2(1), 239-248.
- Field, J. (2008). *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Field, J. (2012). The cognitive validity of the lecture-based question in the IELTS Listening paper. *IELTS Collected Papers*, 2, 391-453.
- Flowerdew, J. (1994). *Academic listening: Research perspectives*. Cambridge: Cambridge University Press.
- Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge: Cambridge University Press.
- Floyd, M., & Jeschull, L. (2012). Teaching with TED Talks: Authentic and motivational language instruction. *The Newsletter of the Video and Digital Media Interest Section*. Retrieved October 08, 2012 from newsmanager.commpartners.com/tesolvdemis/issues/2012-08-10/12.html
- Foster, P., & Skehan, P. (2013). Anticipating a post-task activity: The effects on accuracy, complexity, and fluency of second language performance. *Canadian Modern Language Review*, 69(3), 249-273.

- Frisbie, D. A., & Becker, D. F. (1991). An analysis of textbook advice about true-false tests. *Applied Measurement in Education*, 4, 67-83.
- Gambrell, L. B., Koskinen, P. S., & Kapinus, B. A. (1991). Retelling and the reading comprehension of proficient and less-proficient readers. *The Journal of Educational Research*, 84(6), 356-362.
- Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24(3), 239-258.
- Goh, C. (2000) A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55-75.
- Haastrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen, Germany: Gunter Narr.
- Hernandez, S. S. (2004). *The effects of video and captioned text and the influence of verbal and spatial abilities on second language listening comprehension in a multimedia learning environment*. Unpublished doctoral dissertation, New York University, USA.
- Herron, C. (1994). An investigation of the effectiveness of using an advance organizer to introduce video in the foreign language classroom. *The Modern Language Journal*, 78(2), 190-198.
- Herron, C., York, H., Cole, S. P., & Linden, P. (1998). A comparison study of student retention of foreign language video: Declarative versus interrogative advance organizer. *The Modern Language Journal*, 82(2), 237-247.
- 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, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-30.
- Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement Load Hypothesis versus Technique Feature Analysis. *System*, 56, 28-39.
- Huang, H. C., & Eskey, D. E. (1999-2000). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems*, 28(1), 75-96.

- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long, (Eds.), *The Handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539-558.
- Hwang, Y. (2003). *The effect of the use of video captioning on English as a Foreign Language (EFL) on college students' language learning in Taiwan*. Unpublished doctoral dissertation, The University of Mississippi, USA.
- Hwang, K., & Nation, I. S. P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language*, 6(1), 323-335.
- Imura, H. (2007). The listening process: Effects of question types and repetition. *Language Education & Technology*, 44, 75-85.
- Jafari, K., & Hashim, F. (2012). The effects of using advance organizers on improving EFL learners' listening comprehension: A mixed method study. *System*, 40(2), 270-281.
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?. *Applied Linguistics*, 19(3), 357-377.
- Jones, L. C., & Jan L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86, 546-561.
- Jung, E. H. (2003). The role of discourse signalling cues in second language listening comprehension. *Modern Language Journal*, 87, 562-577.
- Kang, T., Arvizu, M. N. G., Chaipupae, P., & Lesnov, R. O. (2016). Reviews of academic English listening tests for non-native speakers. *International Journal of Listening*, 1-38. doi: 10.1080/10904018.2016.1185210.
- Karakas, A., & Sariçoban, A. (2012). The impact of watching subtitled animated cartoons on incidental vocabulary learning of ELT students. *Teaching English with Technology*, 4, 3-15.
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58(2), 285-325.
- Koolstra, C. M., & Beentjes, J. W. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, 47(1), 51-60.

- Krashen, S. D. (1996). The case for narrow listening. *System*, 24, 97-100.
- Krashen, S. D. (2004). The case for narrow reading. *Language Magazine*, 3(5), 17-19.
- Krashen, S. D. (1980). The input hypothesis. In J. E. Alatis (Ed.), *Georgetown Roundtable on Languages and Linguistics* (pp.168-180). Washington, D.C.: Georgetown University Press.
- Krashen, S. D. (1983). The din in the head, input, and the language acquisition device. *Foreign Language Annals*, 16(1), 41-44.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of multiple true-false items. *Applied Measurement in Education*, 2, 207-216. [doi:10.1207/s15324818ame0203_2](https://doi.org/10.1207/s15324818ame0203_2).
- Kurnia, N. (2003). *Retention of multi-word strings and meaning derivation from L2 reading*. Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.
- Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, 6(1), 10-20.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research*, 21(1), 118-135.
- Leeser, M. J. (2004). The effects of topic familiarity, mode, and pausing on second language learners' comprehension and focus on form. *Studies in Second Language Acquisition*, 26(4), 587-616.
- Lindstromberg, S., & Boers, F. (2008). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, 29(2), 200-222.
- Linacre, J. M. (2007). *Winsteps Rasch measurement* (Version 3.64. 2).
- Linden, M., & Wittrock, M. C. (1981). The teaching of reading comprehension according to the model of generative learning. *Reading Research Quarterly*, 17(1), 44-57.
- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *REL C Journal*, 16(1), 33-42.
- Liversidge, G. B. (2000). *The role of captioning in second language acquisition*. Unpublished doctoral dissertation, Temple University, Philadelphia, USA.

- Long, D. R. (1990). What you don't know can't help you. *Studies in Second Language Acquisition*, 12(01), 65-80.
- Long, M. (1980). *Input, Interaction and Second Language Acquisition*. Unpublished PhD thesis. University of California, Los Angeles, USA.
- Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input1. *Applied Linguistics*, 4(2), 126-141.
- Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 377-393). Rowley, MA: Newbury House.
- Long, M.H. (1996). The role of the linguistics environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of language acquisition: Vol. 4. Second language acquisition* (pp. 413-468). New York: Academic Press.
- Loschky, L. (1994). Comprehensible input and second language acquisition. *Studies in Second Language Acquisition*, 16(03), 303-323.
- Lu, M. (2013). Effects of four vocabulary exercises on facilitating learning vocabulary meaning, form, and use. *TESOL Quarterly*, 47(1), 167-176.
- Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75(2), 196-204.
- Lynch, T. (2009). *Teaching second language listening*. Oxford: Oxford University Press.
- Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 classrooms*. Oxford: Oxford University Press.
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32(3), 321-328.
- Markham, P., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, 34(5), 439-445.
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19, 741-760.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.

- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14, 5-8.
- Morell, T. , García, M., & Sanchez, I. (2008). Multimodal strategies for effective academic presentations in English for non-native speakers. In R. Monroy & A. Sanchez (Eds.), *25 years of applied linguistics in Spain: Milestones and challenges* (pp. 557-568). Universidad de Murcia: Editum.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: 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., & Heatley A. (2002). *Range: A program for the analysis of vocabulary in texts*.
- Retrieved June 10, 2015 from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P., & Newton J. (2009). *Teaching ESL/EFL listening and speaking*. New York: Routledge.
- Nation, I. S. P., & Webb, S. A. (2011). *Researching and analysing vocabulary*. Boston, USA: Heinle, Cengage Learning.
- Neuman, B., & Koskinen P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 96-106.
- Nurmukhamedov, U., & Sadler, R. (2011). Podcasts in Four Categories: Applications to Language Learning. In B. Facer & M. Abdous (Eds.), *Academic podcasting and mobile assisted language learning: Applications and outcomes* (pp. 176-195). Portland, OR: Book News.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55(4), 601-626.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

- Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, 55(4), 174-200.
- Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 09-29.
- Parry, T. S., & Meredith, R. A. (1984). Videotape vs. audiotape for listening comprehension tests: An experiment. *OMLTA journal*, 47, 53.
- Perez, M., Noortgate W., & Desmet P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41, 720-739.
- Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, 53, 233-284.
- Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied linguistics*, 28(1), 66-86.
- Rassaei, E. (2015). Effects of three forms of reading-based output activity on L2 vocabulary learning. *Language Teaching Research*. doi: 1362168815606160.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-90.
- Robinson, P. (1995). Attention, memory, and the “noticing” hypothesis. *Language Learning*, 45, 283-331.
- Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp.631-678). New York: Blackwell Publishing.
- Rodgers, M. (2013) *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.
- Rodgers, M. P., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689-717.
- Rost, M. (2013). *Teaching and researching listening*. New York: Routledge.

- Sadighi, F., & Zare, S. (2006). Is listening comprehension influenced by the background knowledge of the learners? A case study of Iranian EFL learners. *The Linguistics Journal*, 1(3), 110-126.
- Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly*, 21, 221-238.
- Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Quarterly*, 43(2), 360-372.
- Sawada, K. (2009). *Vocabulary acquisition through listening and its relation to learning channel preferences*. Unpublished doctoral dissertation, Temple University, USA.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in Second Language Learning*, 11, 237-326.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu: University of Hawai'i Press.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). New York: Cambridge University Press.
- Schmidt-Rinehart, B. C. (1994). The effects of topic familiarity on second language listening comprehension. *Modern Language Journal*, 78(2), 179-189.
- Schmitt N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.
- Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *TESOL Journal*, 9(1), 4-9.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Seo, K. (2002). Research note: The effect of visuals on listening comprehension: A study of Japanese learners' listening strategies. *International Journal of Listening*, 16, 57-81.
- Shin, D. (1998). Using videotaped lectures for testing academic listening proficiency. *International Journal of Listening*, 12, 57-80.

- Smith Jr, E. V. (2000). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2, 281-311.
- 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.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.
- Stahl, S., & Clark H. (1987). The effects of participatory expectations in classroom discussion on the learning of science vocabulary. *American Educational Research Journal*, 24(4), 541-555.
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.
- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661-699.
- Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). New York: Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & G. Seidhofer (Eds.), *Principles and practices in applied linguistics: Studies in honour of H. G. Widdowson*, (pp. 125-144). Oxford: Oxford University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-83). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50-73.
- Takaesu, A. (2013). TED Talks as an extensive listening resource for EAP students. *Language Education in Asia*, 4, 150-162.
- van Zeeland, H. (2013). Vocabulary and listening. In C. A. Chapelle (Ed.), *The Encyclopaedia of Applied Linguistics* (pp.01-06). Australia: Wiley-Blackwell.

- van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34, 457-479.
- van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41(3), 609-624.
- Vandergrift, L. (2004). Listening to learn or learning to listen?. *Annual Review of Applied Linguistics*, 24, 3-25.
- Vandergrift, L., & Goh, C. C. (2012). *Teaching and learning second language listening: Metacognition in action*. Oxford: Taylor & Francis Group.
- Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, 60(2), 470-497.
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition?. *Applied Linguistics*, 24(1), 56-89.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258.
- Wagner, E. (2002). Video listening tests: A pilot study. *Working Papers in TESOL and Applied Linguistics*, 2, 1-39.
- Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, 27, 493-513.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65.
- Webb, S. 2014. Repetition in incidental vocabulary learning. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–6). New York: John Wiley and Sons.
- Webb, S., & Chang, A. C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, 68(3), 267-290.
- Webb, S., & Chang, A. C. (2014). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning?. *Language Teaching Research*. doi:1362168814559800.
- Webb, S., & Chang, A. C. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651-675.

- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.
- Wilson, R. M., Gambrell L. B., & Pfeiffer, W. R. (1985). The effects of retelling upon reading comprehension and recall of text information. *The Journal of Educational Research*, 78(4), 216-220.
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65-86.
- Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1), 254-275.
- Wittrock, M. C. (1991). Generative Teaching of Comprehension. *Elementary School Journal*, 92(2), 169-84.
- Wittrock, M. C., & Kelly, R. (1984). *Teaching reading comprehension to adults in basic skills courses* (Report MDA 903-82-C-0169). Graduate School of Education, University of California, Los Angeles.
- Yi, B., & Sun, Z. (2013). An empirical study of the effectiveness of negotiation of meaning in L2 vocabulary acquisition of Chinese learners of English. *English Language Teaching*, 6(10), 120-131.
- Yüksel, D., & Tanriverdi B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *The Turkish Online Journal of Educational Technology*, 8(2), 48–54.
- Zidong, H. (1998). The effects of topic familiarity, proficiency level and question type on EFL listening comprehension: Schemata Theory and Relevance Theory perspectives [J]. *Morden Foreign Languages*, 4.

APPENDIXES

A. Information sheet

SCHOOL OF LINGUISTICS & APPLIED LANGUAGE STUDIES

TE KURA TATARI REO

Level 2, Von Zedlitz Building, Kelburn Parade,

PO Box 600 Wellington 6140

Phone +64 4 463 5600 Fax +64 4 463 5604 Email lals@vuw.ac.nz Website www.victoria.ac.nz/lals

Information Sheet

Research Project: *Investigating the merits of TED Talks as English as a Foreign Language (EFL) listening materials*

Researcher: *Nguyen, Chi Duc, School of Linguistics and Applied Language Studies, Victoria University of Wellington*

I am a PhD student in Applied Linguistics at Victoria University of Wellington. As part of this degree, I am undertaking a research project leading to a thesis. The project I am undertaking is examining different merits of TED Talks as EFL listening materials. This research project has received approval from the Victoria University Human Ethics Committee.

I am now inviting Vietnamese learners of EFL who have an English language proficiency level equivalent to overall IELTS band of 5.5 to participate in this study. You will be asked to complete a 30-minute vocabulary size test, two 50-minute watching sessions of selected TED talks, a 10-minute comprehension test, and a 5-minute follow-up interview or questionnaire survey. Your interaction during the experiment will be audio-recorded. All of these will be run in a language classroom in the Faculty of English Language Teachers Education, University of Languages and International Studies, Vietnam National University of Hanoi, 1 Pham Van Dong Street, Cau Giay, Hanoi, Vietnam, at a time that is convenient to you.

Participating in this research project is voluntary. It has no impact on your classes and the results do not affect your course grades in any way. Additionally, you can withdraw from the project at any time before December 15, 2015 without having to give any reasons. As a token of appreciation for your participation in this research project, you will receive a small present, your score on the L2 Vocabulary Size Test, and practical advice on L2 vocabulary learning. You will also be provided with the test results related to your L2 listening comprehension and incidental vocabulary uptake through watching the assigned TED talks. If you indicate on the consent form that you wish to be informed of the general outcome of the study, a summary of that outcome will be sent to you via e-mail.

Your participation will form the basis of my research project and will be put into a written report on an anonymous basis. It will not be possible for you to be identified personally. All materials collected will be kept confidential. No other person besides me and my supervisors, Associate Professor Frank Boers and Dr. Jonathan Newton, and authorized transcribers will have access to those materials. The thesis will be submitted for marking to the School of Linguistics and Applied Language Studies, Victoria University of Wellington, and then deposited in the University Library. It is intended that one or more articles will be submitted for publication in scholarly journals. All the materials will be destroyed immediately at the end of the research project.

If you have any further questions or would like to receive further information about the project, please contact me at Chiduc.nguyen@vuw.ac.nz, or my primary supervisor Associate Professor Frank Boers, the School of Linguistics and Applied Language Studies at Victoria University of Wellington, at Frank.Boers@vuw.ac.nz, or via his phone number 04 4636014.

Nguyen, Chi Duc, Email: Chiduc.nguyen@vuw.ac.nz Mobile: 0936777206

School of Linguistics and Applied Language Studies, Victoria University of Wellington

B. Consent form (a signed sample)

<p>SCHOOL OF LINGUISTICS & APPLIED LANGUAGE STUDIES TE KURA TATARI REO Level 2, Von Zedlitz Building, Kelburn Parade, PO Box 600 Wellington 6140 Phone +64 4 463 5600 Fax +64 4 463 5604 Email lals@vuw.ac.nz Website www.victoria.ac.nz/lals</p>
---	---

Consent Form

Consent to participate in:

Research Project: *Investigating the merits of TED Talks as English as a Foreign Language (EFL) listening materials*

Researcher: *Nguyen, Chi Duc, School of Linguistics and Applied Language Studies, Victoria University of Wellington*

I have been given and have understood an explanation of this research project. I have had an opportunity to ask questions and have them answered to my satisfaction. I understand that I may withdraw myself (or any information I have provided) from this project before December 15, 2015 without having to give reasons.

I understand that any information I provide will be kept confidential to the researcher, the supervisors and the person who transcribes the tape recordings of my interaction with other participants during the experiment. I understand the published results will not use my name, and that no opinions will be attributed to me in any way that will identify me. I understand that the tape recordings of our interaction will be wiped immediately after the research project has been completed unless I indicate that I would like them to be returned to me. I also understand that my participation in this project has no impact on my classes and the results do not affect my course grades in any way.

	Yes or No
I would like to receive a summary of the results of the research when it is completed.	Yes
I agree to take part in this research.	Yes

Signed:

Name of participant:

Date: 10/11/2015

C. Profiles of research participants

Control condition (Code: CT0)

Code	General information		English language proficiency	
	Age	Gender	L2 Proficiency Scores	Vocabulary Levels Test Scores
CT001	19	Female	8	5,700
CT002	20	Female	8.1	5,400
CT003	20	Female	7.8	4,000
CT004	20	Female	7.8	4,000
CT005	20	Male	8.3	4,600
CT006	20	Female	8.7	5,300
CT007	20	Female	8.2	5,700
CT008	20	Female	7.7	4,000
CT009	20	Female	7.8	4,000
CT010	20	Female	8.5	5,600
CT011	20	Female	8.7	5,600
CT012	20	Female	8.6	5,400
CT013	20	Female	7.7	4,000
CT014	19	Female	8.6	5,700
CT015	20	Female	8.3	5,300
CT016	20	Female	8.9	4,900
CT017	20	Female	8.3	4,000
CT018	20	Female	8.4	5,900
CT019	20	Female	8.3	5,900
CT020	20	Female	8.3	4,000
CT021	20	Female	8.1	4,000
CT022	20	Female	8.3	5,500
CT023	20	Female	8.2	5,200
CT024	20	Female	8.5	4,000
CT025	20	Female	8.3	4,000
CT026	20	Male	8.3	4,000
CT027	20	Female	8.3	4,000
CT028	20	Female	8.3	5,000
CT029	20	Female	9	6,100
CT030	20	Female	7.5	4,000
CT031	20	Female	8.3	4,400
CT032	20	Female	8.6	4,400

Note: Maximum L2 proficiency score: 10; Maximum VST score: 14,000

OV condition (Code: OV1)

Code	General information		English language proficiency	
	Age	Gender	L2 Proficiency Scores	Vocabulary Levels Test Scores
OV101	20	Female	8.1	4400
OV102	20	Female	8.2	4000
OV103	20	Female	8.8	5600
OV104	20	Female	8.3	4600
OV105	20	Female	8.1	4300
OV106	20	Female	8.2	4900
OV107	20	Female	8.4	4100
OV108	20	Female	8.1	4000
OV109	20	Female	8.3	4900
OV110	20	Female	8.2	4600
OV111	20	Female	8	4700
OV112	20	Female	8.4	6200
OV113	20	Female	8.1	6100
OV114	20	Female	8.1	4000
OV115	20	Female	8.6	5500
OV116	20	Female	7.5	5100
OV117	20	Female	8.6	4700
OV118	20	Female	8.9	5000
OV119	20	Female	8.4	5600
OV120	20	Female	7.6	4300
OV121	19	Female	8.2	4100
OV122	20	Female	8.6	4000
OV123	20	Female	8.5	4500
OV124	20	Female	8.1	4000
OV125	20	Female	8.3	4800
OV126	20	Female	7.7	4000
OV127	20	Female	8.1	4000
OV128	20	Female	8	4000
OV129	20	Female	8.1	4000
OV130	20	Female	8	4800
OV131	20	Female	7.7	4900
OV132	20	Female	8.5	7000

Note: Maximum L2 proficiency score: 10; Maximum VST score: 14,000

RV condition (Code: RV2)

Code	General information		English language proficiency	
	Age	Gender	L2 Proficiency Scores	Vocabulary Levels Test Scores
RV201	20	Female	8.7	4900
RV202	20	Female	8.4	7100
RV203	20	Female	8.4	5700
RV204	20	Female	8.6	4100
RV205	20	Female	8.5	6400
RV206	20	Female	8.1	6200
RV207	20	Female	8.9	5000
RV208	20	Female	8.5	4600
RV209	20	Female	8.5	6400
RV210	20	Female	8.7	4400
RV211	20	Female	8.2	4000
RV212	20	Female	8.3	6700
RV213	20	Female	8	4800
RV214	20	Female	8.8	5600
RV215	20	Female	8.1	5300
RV216	20	Female	7.9	5200
RV217	20	Female	8.4	4000
RV218	20	Female	8.8	5800
RV219	20	Female	8.3	4000
RV220	20	Female	8.1	4600
RV221	20	Female	7.9	4000
RV222	20	Female	7.6	8000
RV223	20	Female	8.6	4900
RV224	20	Female	7.9	4000
RV225	20	Female	8.7	4000
RV226	20	Female	8.2	4100
RV227	20	Female	8.7	4000
RV228	20	Female	8.5	4900
RV229	20	Female	8.6	4100
RV230	20	Male	8.2	4000
RV231	20	Female	7.9	5000
RV232	20	Female	8.2	4000

Note: Maximum L2 proficiency score: 10; Maximum VST score: 14,000

RVS condition (Code: RVS3)

Code	General information		English language proficiency	
	Age	Gender	L2 Proficiency Scores	Vocabulary Levels Test Scores
RVS301	20	Female	7.5	4000
RVS302	20	Female	8.8	5400
RVS303	20	Female	8.2	4000
RVS304	20	Female	8.3	4600
RVS305	20	Male	8.7	5300
RVS306	20	Female	7.1	5000
RVS307	20	Female	8.9	6100
RVS308	20	Female	8.3	5500
RVS309	20	Female	7.8	5200
RVS310	20	Female	7.3	4000
RVS311	20	Female	8.1	4800
RVS312	20	Female	7.5	5600
RVS313	20	Female	8	4800
RVS314	19	Female	7.5	5300
RVS315	20	Female	8.7	4700
RVS316	20	Female	8.1	4600
RVS317	20	Female	8.5	4800
RVS318	20	Female	8.7	5500
RVS319	20	Female	9	5800
RVS320	20	Female	8.9	6100
RVS321	20	Female	8.3	5800
RVS322	20	Female	8	5600
RVS323	20	Female	9	5900
RVS324	20	Female	9.1	4800
RVS325	20	Female	8.3	4000
RVS326	20	Female	9.3	5500
RVS327	20	Female	8.4	5700
RVS328	20	Female	8.9	5100
RVS329	20	Female	8.4	4600
RVS330	20	Female	8.2	4000
RVS331	20	Female	8.4	5300
RVS332	20	Female	7.8	4800

Note: Maximum L2 proficiency score: 10; Maximum VST score: 14,000

NV condition (Code: NV4)

Code	General information		English language proficiency	
	Age	Gender	L2 Proficiency Scores	Vocabulary Levels Test Scores
NV401	20	Female	8.1	4400
NV402	20	Female	8.2	4000
NV403	20	Female	8.8	5600
NV404	20	Female	8.3	4600
NV405	20	Female	8.1	4300
NV406	20	Female	8.2	4900
NV407	20	Female	8.4	4100
NV408	20	Female	8.1	4000
NV409	20	Female	8.3	4900
NV410	20	Female	8.2	4600
NV411	20	Female	8	4700
NV412	20	Female	8.4	6200
NV413	20	Female	8.1	6100
NV414	20	Female	8.1	4000
NV415	20	Female	8.6	5500
NV416	20	Female	7.5	5100
NV417	20	Female	8.6	4700
NV418	20	Female	8.9	5000
NV419	20	Female	8.4	5600
NV420	20	Female	7.6	4300
NV421	20	Female	8.2	4100
NV422	20	Female	8.6	4000
NV423	20	Female	8.5	4500
NV424	20	Female	8.1	4000
NV425	20	Female	8.3	4800
NV426	20	Female	7.7	4000
NV427	20	Female	8.1	4000
NV428	20	Female	8	4000
NV429	20	Female	8.1	4000
NV430	20	Female	8	4800
NV431	20	Female	7.7	4900
NV432	20	Female	8.5	7000

Note: Maximum L2 proficiency score: 10; Maximum VST score: 14,000

NVS condition (Code: NVS5)

Code	General information		English language proficiency	
	Age	Gender	L2 Proficiency Scores	Vocabulary Levels Test Scores
NVS501	20	Female	7.1	4000
NVS502	20	Female	7.5	4000
NVS503	20	Female	6.8	4000
NVS504	20	Female	6.9	4000
NVS505	20	Female	7.6	4100
NVS506	20	Female	8	4900
NVS507	20	Female	8.2	5100
NVS508	20	Female	8.2	4000
NVS509	20	Male	8.4	4600
NVS510	19	Female	8.8	5200
NVS511	20	Female	8.1	4000
NVS512	20	Female	8.8	4300
NVS513	20	Female	8.6	4000
NVS514	20	Female	8.9	4000
NVS515	20	Female	8.1	4600
NVS516	20	Female	8.8	4600
NVS517	20	Female	8.2	4200
NVS518	20	Female	8.3	6700
NVS519	20	Female	8.3	4800
NVS520	20	Female	8.4	4000
NVS521	20	Female	8.5	5100
NVS522	20	Female	8.4	5000
NVS523	19	Female	8.6	6300
NVS524	20	Female	8.3	5400
NVS525	20	Female	8.2	6400
NVS526	20	Female	8.2	7200
NVS527	20	Female	8.8	7100
NVS528	20	Female	8.7	6200
NVS529	20	Female	8.6	6800
NVS530	20	Female	8.2	5300
NVS531	20	Female	8.3	4900
NVS532	20	Male	8.8	7100

Note: Maximum L2 proficiency score: 10; Maximum VST score: 14,000

Study 7 (Code: RVS6 and NVS6)

Code	General information		English language proficiency	
	Age	Gender	IELTS Scores	Vocabulary Levels Test Scores
RVS601	33	Female	6.5	6000
RVS602	33	Female	7	7200
RVS603	31	Female	7.5	6500
RVS604	30	Female	7.5	7100
NVS601	36	Female	6.5	7500
NVS602	34	Female	7	6900
NVS603	27	Female	7.5	7400
NVS604	29	Female	7.5	7300

D. Information on experimental TED Talks

Video 1: https://www.ted.com/playlists/7/ocean_wonders

Video 2: https://www.ted.com/talks/deep_ocean_mysteries_and_wonders

Video 3: https://www.ted.com/talks/edith_widder_the_weird_and_wonderful_world_of_bioluminescence

Titles	Presenters	Length
Underwater astonishments	David Gallo	5:00
Deep ocean mysteries and wonders	David Gallo	8:00
The weird and wonderful world of bioluminescence	Edith Widder	12:00

Note: The length excludes the opening and ending footage (about 30 seconds) that presents general information about TED Talks.

Transcript 1:

We are going to go on a dive to the deep sea, and anyone that has had that lovely opportunity knows that for about two and half hours on the way down, it is a perfectly positively pitch-black world. And we used to see the most mysterious animals out the window that you could

not describe: these blinking lights -- a world of bioluminescence, like fireflies. Dr. Edith Widder -- she is now at the Ocean Research and Conservation Association -- was able to come up with a camera that could capture some of these incredible animals, and that is what you are seeing here on the screen.

That is all bioluminescence. So, like I said: just like fireflies. There is a flying turkey under a tree. I am a geologist by training. But I love that. And you see, some of the bioluminescence they use to avoid being eaten, some they use to attract prey, but all of it, from an artistic point of view, is positively amazing. And a lot of what goes on inside. There is a fish with glowing eyes, pulsating eyes. Some of the colours are designed to hypnotize, these lovely patterns. And then this last one, one of my favourites, this pinwheel design. Just absolutely amazing, every single dive.

That is the unknown world, and today we have only explored about 3 percent of what is out there in the ocean. Already we have found the world's highest mountains, the world's deepest valleys, underwater lakes, underwater waterfalls a lot of that we shared with you from the stage. And in a place where we thought no life at all, we find more life, we think, and diversity and density than the tropical rainforest, which tells us that we do not know much about this planet at all. There is still 97 percent, and either that 97 percent is empty or just full of surprises.

But I want to jump up to shallow water now and look at some creatures that are positively amazing. Cephalopods head foos. As a kid I knew them as calamari, mostly. This is an octopus this is the work of Dr. Roger Hanlon at the Marine Biological Lab and it is just fascinating how cephalopods can, with their incredible eyes, sense their surroundings, look at

light, look at patterns. Here is an octopus moving across the reef, finds a spot to settle down, curls up and then disappears into the background. Tough thing to do.

In the next bit, we are going to see a couple squid. These are squid. Now males, when they fight, if they are really aggressive, they turn white. And these two males are fighting. They do it by bouncing their butts together, which is an interesting concept. Now, here is a male on the left and a female on the right, and the male has managed to split his coloration so the female only always sees the kinder gentler squid in him. And the male. We are going to see it again. Let us take a look at it again. Watch the coloration: white on the right, brown on the left. He takes a step back so he is keeping off the other males by splitting his body and comes up on the other side. Bingo! Now I am told that is not just a squid phenomenon with males, but I do not know.

Cuttlefish. I love cuttlefish. This is a Giant Australian Cuttlefish. And there he is, his droopy little eyes up here. But they can do pretty amazing things, too. Here we are going to see one backing into a crevice, and watch his tentacles he just pulls them in, makes them look just like algae. Disappears right into the background. Positively amazing. Here is two males fighting. Once again, they are smart enough, these cephalopods; they know not to hurt each other. But look at the patterns that they can do with their skin. That is an amazing thing.

Here is an octopus. Sometimes they do not want to be seen when they move because predators can see them. Here, this guy actually can make himself look like a rock, and, looking at his environment, can actually slide across the bottom, using the waves and the shadows so he cannot be seen. His motion blends right into the background the moving rock trick. So, we are learning lots new from the shallow water. Still exploring the deep, but learning lots from the shallow water. There is a good reason why the shallow water is full of

predators. Here is a barracuda and if you are an octopus or a cephalopod, you need to really understand how to use your surroundings to hide.

In the next scene, you are going to see a nice coral bottom. And you see that an octopus would stand out very easily there if you could not use your camouflage, use your skin to change colour and texture. Here is some algae in the foreground and an octopus. Is not that amazing? Now, Roger spooked him so he took off in a cloud of ink, and when he lands, the octopus says: “Oh, I have been seen. The best thing to do is to get as big as I can get”. That big brown makes his eyespot very big. So, he is bluffing. Let us do it backwards I thought he was joking when he first showed it to me. I thought it was all graphics so here it is in reverse. Watch the skin colour; watch the skin texture. Just an amazing animal, it can change colour and texture to match the surroundings. Watch him blend right into this algae. One, two, three. And now he is gone, and so am I. Thank you very much.

Transcript 2:

You know, I had a real rough time in school with ADD, and I have a PhD. I earned a PhD, but it is tough to pay attention to biology, geology, physics, chemistry really tough for me. Only one thing grabbed my attention. And it is that planet called Earth, but in this picture right here you will see that Earth is mostly water that is the Pacific. 70 percent of Earth is covered with water and you could say: “Hey I know planet Earth. I live here”. You do not know Earth.

You do not know this planet because most of it is covered with that average depth is 2 miles. And when you go outside and look up at the like the Empire State building, Chrysler building, the average depth of the ocean is 15 of those on top of one another. We have explored about 5 percent of what is in that water. Explored, meaning for the first time go

peek and see what is there. So what I want to do today is I want to show you some things about this planet, about the oceans. I want to take you from some shallow water down to the deep water, and hopefully, like me, you will see some things that get you hooked on exploring planet Earth.

You know things like corals, you have seen plenty of corals, those of you that have been to the beach, snorkel, you know corals are amazing places to go full of life, some big animals, small animals, some nice, some dangerous, sharks, whales, all that stuff. They need to be protected from humanity. They are great places, but what you probably do not know about is in the deep ocean, the very deep part of the ocean, we have volcanic eruptions. Most volcanoes on Earth are at the bottom of the sea, more than 80 percent and we actually have fire, fire deep inside the ocean, going on right now. All over the world, in the Pacific, the Atlantic, the Indian Ocean, in this place, the ocean floor, the rocks actually turn to liquid. So you actually have waves on the ocean floor. You would say nothing could live there, but when we look in detail, even there, even in the deepest, darkest places on earth, we find life, which tells us that life really wants to happen. So, pretty amazing stuff.

Every time we go to the bottom of the sea, we explore with our submarines, with our robots, we see something that is usually surprising, sometimes it is startling and sometimes revolutionary. You see that puddle of water sitting there. And all around the water there is a little cliff. There is a little white sandy beach. We will get closer to it. You will see the beach a little bit better, some of the waves in that water, down there. The thing that is special about this water is that it is at the bottom of the Gulf of Mexico. So you are sitting inside of submarine looking out the window at a little pond of water beneath the sea. You know we see ponds, we see lakes, we see rivers in fact right here is a river at the bottom of the ocean going from the lower left to the upper right. Water is actually flowing through there. This totally

blew our minds: how can you have this at the bottom? You are in the ocean looking at more water. And there is animals that only live in that water.

So, the bottom of the ocean. I love this map because it shows in the middle of the ocean there is a mountain range. That mountain range is the greatest mountain range on Earth. It is called the Mid-Ocean Ridge 50,000 miles long, and we have hardly had a peek at it. Hardly had a peek at it. We find valleys, many thousands of valleys larger, wider, deeper than the Grand Canyon. We find, as I said, underwater lakes, rivers, waterfalls. The largest waterfall on the planet is actually under the ocean up near Iceland. All that stuff in that 5 percent that we have explored.

So the deal about the ocean is that to explore it you have got to have technology. Not only technology, but it is not just Dave Gallo or one person exploring. It is a team of people. You have got to have the talent. You have got to have the team. You have got to have the technology and in this case it is our ship, Atlantis, and the submarine, Alvin. Now, inside that submarine this is an Alvin launch there is three people. They are being wheeled out onto deck. There is 47 other people. The team work on that ship making sure that these people are okay. Everybody on that submarine is thinking one thing right now: should I have gone to the bathroom one more time. Because you are in there for ten hours ten hours in that little sphere. Three of you together and nobody is going to be around you. You go into the water and once you hit the water it is amazing. There is a lovely colour blue that penetrates right inside you. You do not hear the surface ship anymore. You hear that pinging of a sonar. If you have got an iPhone you have got sonar on there it is that same pinging that goes down to the bottom and comes back up. Divers check out the sub to make sure the outside is okay, and then they say: "Go", and down you go to the bottom of the ocean and it is an amazing trip. So for two and a half hours you sink down to the bottom. And two hours of it is totally pitch black. And

now we thought that nothing could live inside that world at the bottom of the ocean. And when we look, we find some amazing things.

All the way down we call it the mid-water from the top of the ocean down to the bottom we find life. Whenever we stop and look we find life. I am going to show you some jellies here because they are absolutely some of the coolest creatures on Earth. Look at that thing just flailing his arms around. That is like a little lobster. That one is like all these animals with their mouths hooked together. They are colonial animals. Some animals are tiny. Some can be longer than this stage. Just amazing animals and you cannot collect them with a net. We have to go there with our cameras and take a look at them. So every time we go, new species of life. The ocean is full of life.

And yet the deepest part of the ocean when we go to that mountain range, we find hot springs. Now we were sure, because this is poisonous water, because it is so deep that it will crush the Titanic the same way you crush an empty cup inside your hand. We were sure there would be no life there at all. Instead we find more life and diversity and density than the tropical rainforest. So, in one instance, in one peek out the window of the sub, we discover something that revolutionizes the way we think about life on Earth, and that is you do not always have to have sunlight to get life going. There is big animals down there too, some that look familiar. That guy is called Dumbo. I love him. Dumbo is great. This guy, oh man, I wish I had more footage of this. We are trying to get an expedition together to go look at this and maybe in a year we will have that. Go online and look: *Vampyroteuthis infernalis*, the Vampire Squid. Incredibly cool. In the darkness of the deep sea he is got glowing tentacles, so if I am coming at you like him, I put my arms out in the, the darkness so all you see are these little glowing things over here. In the meantime, I am coming at you. When he wants to escape, he has got these glowing pods on his butt that look like eyes. He has got glowing eyes

on his butt. How cool is that? It is just an amazing, amazing animal. Vampire squid, when it gets protective, it pulls this black cape over its whole body, curls up into a ball. Outrageous animal.

This ship, the ship of dreams a hundred years ago. This coming April, this ship was supposed to show up in New York. It is the Titanic and I co-lead an expedition out there last year. We are learning so much about that ship. Titanic is an interesting place for biology because animals are moving in to live on the Titanic. Microbes are actually eating the hull of Titanic. That is where Jack was king of the world right there on the bow of Titanic. So, we are doing real good and what is exciting me is that someday we are making a virtual Titanic so you can sit there at home with your joystick and your headset on, and you can actually explore Titanic for yourself. That is what we want to do make these virtual worlds so it is not Dave Gallo or someone else exploring the world; it is you. You explore it for yourself.

So here is the bottom line. The oceans are unexplored and I cannot begin to tell you how important that is because they are important to us. Seven billion people live on this planet and all of us are impacted by the sea because the oceans control the air you breathe, the water you drink, the food you eat. All those are controlled in some way by the ocean and this is a thing that we haven't even explored five percent. The thing I want to leave you with is in that 5 percent I showed you some cool stuff. There is a lot more cool stuff every dive we go out on in the ocean, we find something new about the sea. So what is in that other 95 percent? Did we get the exciting stuff or is there more out there? And I am here to tell you that the ocean is full of surprises. There is a quote I love by Marcel Proust: The true voyage of exploration is not so much in seeking new landscapes, which we do, but in having new eyes. And so I hope today by showing you some of this, it is given you some new eyes about this planet, and for the first time I want you to think about it differently. Thank you very much. Thank you.

Transcript 3:

So I want to take you on a trip to an alien world. And it is not a trip that requires light years of travel, but it is to a place where it is defined by light.

So it is a little appreciated fact that most of the animals in our ocean make light. I have spent most of my career studying this phenomenon called bioluminescence. I study it because I think understanding it is critical to understanding life in the ocean where most bioluminescence occurs. I also use it as a tool for visualizing and tracking pollution. But mostly I am entranced by it. Since my, my first dive in a deep diving submersible, when I went down and turned out the lights and saw the fireworks displays, I have been a bioluminescence junky. But I would come back from those dives and try to share the experience with words, and they were totally inadequate to the task. I needed some way to share the experience directly. And the first time I figured out that way was in this little single person submersible called Deep Rover.

This next video clip, you are going to see how we stimulated the bioluminescence. And the first thing you are going to see is a transect screen that is about a meter across.

In front of the sub, a mesh screen will come into contact with the soft bodied creatures of the deep sea. With the sub's lights switched off, it is possible to see their bioluminescence the light produced when they collide with the mesh. This is the first time it has ever been recorded.

So I recorded that with an intensified video camera that has about the sensitivity of the fully dark adapted human eye. Which means that really is what you would see if you took a dive in

a submersible. But just to try to prove that fact to you, I have brought along some bioluminescent plankton in what is undoubtedly a foolhardy attempt at a live demonstration.

So, if we could have the lights down and have it as dark in here as possible, I have a flask that has bioluminescent plankton in it. And you will note there is no light coming from them right now, either because they are dead or because I need to stir them up in some way for you to see what bioluminescence really looks like. Oops. Sorry. I spend most of my time working in the dark; I am used to that. Okay.

So that light was made by a bioluminescent dinoflagellate, a single celled alga. So why would a single-celled alga need to be able to produce light? Well, it uses it to defend itself from its predators. The flash is like a scream for help. It is what is known as a bioluminescent burglar alarm, and just like the alarm on your car or your house, it is meant to cast unwanted attention onto the intruder, thereby either leading to his capture or scaring him away.

There is a lot of animals that use this trick, for example this black dragonfish. It has got a light organ under its eye. It has got a chin barbell. It has got a lot of other light organs you cannot see, but you will see in here in a minute. So we had to chase this in the submersible for quite some time, because the top speed of this fish is one knot, which was the top speed of the submersible. But it was worth it, because we caught it in a special capture device, brought it up into the lab on the ship, and then everything on this fish lights up. It is unbelievable. The light organs under the eyes are flashing. That chin barbell is flashing. It has got light organs on its belly that are flashing, fin lights. It is a scream for help; it is meant to attract attention. It is phenomenal. And you normally do not get to see this because we have exhausted the luminescence when we bring them up in nets.

There are other ways you can defend yourself with light. For example, this shrimp releases its bioluminescent chemicals into the water just the way a squid or an octopus would release an ink cloud. This blinds or distracts the predator. This little squid is called the fire shooter because of its ability to do this. Now it may look like a tasty morsel, or a pig's head with wings but if it is attacked, it puts out a barrage of light in fact, a barrage of photon torpedoes. I just barely got the lights out in time for you to be able to see those gobs of light hitting the transect screen and then just glowing. It is phenomenal.

So there is a lot of animals in the open ocean most of them that make light. And we have a pretty good idea, for most of them, why. They use it for finding food, for attracting mates, for defending against predators. But when you get down to the bottom of the ocean that is where things get really strange. And some of these animals are probably inspiration for the things you saw in Avatar, but you do not have to travel to Pandora to see them. They are things like this.

This is a golden coral, a bush. It grows very slowly. In fact, it is thought that some of these are as much as 3,000 years old, which is one reason that bottom trawling should not be allowed. The other reason is this amazing bush glows. So if you brush up against it, any place you brushed against it, you get this twinkling blue green light that is just breathtaking. And you see things like this. This looks like something out of a Dr. Seuss book just all manner of creatures all over this thing. And these are flytrap anemones. Now if you poke it, it pulls in its tentacles. But if you keep poking it, it starts to produce light. And it actually ends up looking like a galaxy. It produces these strings of light, presumably as some form of defence.

There are starfish that can make light. And there are brittle stars that produce bands of light that dance along their arms. This looks like a plant, but it is actually an animal. And it

anchors itself in the sand by blowing up a balloon on the end of its stock. So it can actually hold itself in very strong currents, as you see here. But if we collect it very gently, and we bring it up into the lab and just squeeze it at the base of the stock, it produces this light that propagates from stem to the plume, changing colour as it goes, from green to blue. Colorization and sound effects added for your viewing pleasure. But we have no idea why it does that.

Here is another one. This is also a sea pen. It has got a brittle star hitching a ride. It is a green saber of light. And like the one you just saw, it can produce these as bands of light. So if I squeeze the base, the bands go from base to tip. If I squeeze the tip, they go from tip to base. So what do you think happens if you squeeze it in the middle, I had been very interested in your theories about what that is about.

So there is a language of light in the deep ocean, and we are just beginning to understand it, and one way we are going about that is we are imitating a lot of these displays. This is an optical lure that I have used. We call it the electronic jellyfish. It is just 16 blue LEDs that we can program to do different types of displays. And we view it with a camera system I developed called Eye in the Sea that uses far red light that is invisible to most animals, so it is unobtrusive. So I just want to show you some of the responses we have elicited from animals in the deep sea.

So the camera's black and white. It is not high resolution. And what you are seeing here is a bait box with a bunch of like the cockroaches of the ocean there are isopods all over it. And right in the front is the electronic jellyfish. And when it starts flashing, it is just going to be one of the LEDs that is flashing very fast. But as soon as it starts to flash and it is going to look big, because it blooms on the camera I want you to look right here. There is something

small there that responds. We are talking to something. It looks like a little of string pearls basically in fact, three strings of pearls. And this was very consistent. This was in the Bahamas at about 2,000 feet. We basically have a chat room going on here, because once it gets started, everybody is talking. And I think this is actually a shrimp that is releasing its bioluminescent chemicals into the water. But the cool thing is we are talking to it. We do not know what we are saying. Personally, I think it is something sexy.

And then finally, I want to show you some responses that we recorded with the world's first deep sea webcam, which we had installed in Monterey Canyon last year. We have only just begun to analyze all of this data. This is going to be a glowing source first, which is like bioluminescent bacteria. And it is an optical cue that there is carrion on the bottom of the ocean. So this scavenger comes in, which is a giant sixgill shark. And I cannot claim for sure that the optical source brought it in, because there is bait right there. But if it had been following the odour plume, it would have come in from the other direction. And it does actually seem to be trying to eat the electronic jellyfish. That is a 12 foot long giant sixgill shark.

Okay, so this next one is from the webcam, and it is going to be this pinwheel display. And this is a burglar alarm. And that was a Humboldt squid, a juvenile Humboldt squid, about three feet long. This is at 3,000 feet in Monterey Canyon. But if it is a burglar alarm, you wouldn't expect it to attack the jellyfish directly. It is supposed to be attacking what is attacking the jellyfish. But we did see a bunch of responses like this. This guy is a little more contemplative. Hey, wait a minute. There is supposed to be something else there. He is thinking about it. But he is persistent. He keeps coming back. And then he goes away for a few seconds to think about it some more, and thinks: "Maybe if I come in from a different angle". Nope.

So we are starting to get a handle on this, but only just the beginnings. We need more eyes on the process. So if any of you ever get a chance to take a dive in a submersible, by all means, climb in and take the plunge. This is something that should be on everybody's bucket list, because we live on an ocean planet. More than 90 percent, 99 percent, of the living space on our planet is ocean. It is a magical place filled with breathtaking light shows and bizarre and wondrous creatures, alien life forms that you do not have to travel to another planet to see. But if you do take the plunge, please remember to turn out the lights. But I warn you, it is addictive. Thank you.

E. Word-form recall test

Vocabulary Test 1.2 (PRT)

Full Name:

Class:

Age:

Gender:

INSTRUCTION (Hướng dẫn):

Trong bài kiểm tra này, các em sẽ cố gắng nhớ lại **dạng thức của 40 từ đơn lẻ** từng **xuất hiện trong video** các em vừa xem. Để nhớ mỗi dạng thức từ, các em sẽ được cung cấp **2 dữ liệu**: (1) **nét nghĩa Tiếng Việt** của từ và (2) **âm tiết đầu** của từ đó. Trong **20 giây**, các em cố gắng nhớ lại dạng thức của từ đó và **nói ra từ đó với máy ghi âm**. Chú ý: Các em có thể đưa ra **nhiều hơn một đáp án** cho mỗi dạng thức từ. Nếu **chưa nhớ** được dạng thức từ đó, các em nói "**no response**". Dưới đây là một **ví dụ**:

Trong phiếu trả lời, các em sẽ thấy:

1. tàu ngầm.

Trong bảng, các em sẽ nghe:

1. Sub

Với máy ghi âm, các em sẽ nói:

1. Submarine HOẶC 1. No response

01	vỏ tàu	21	ma cà rồng
02	phóng ra, truyền dẫn	22	cần điều khiển từ xa
03	lặn xuống	23	râu (cá)
04	mắt lưới	24	luồng (ánh sáng)
05	Mỗi	25	không gây ảnh hưởng
06	đẹp sững sốt	26	(mắt) xệ
07	miếng (ăn)	27	vết sáng như thanh đao
08	phát quang sinh học	28	thôi mien
09	ngoại lai, xâm chiếm	29	trêu trọc, dọa nạt
10	thước phim	30	dao động
11	lưới quét	31	bị mê hoặc
12	truyền âm	32	mồi giả
13	vết nứt	33	lừa phỉnh
14	vẫy gọi	34	dại đột
15	phiêu sinh vật	35	ngụy trang
16	đom đóm	36	kỳ lạ, thú vị
17	lấp lánh	37	lặn với ống thở
18	xem xét, quan sát	38	đẹp đến ngạc nhiên
19	lông, xúc tu nhỏ	39	(người) nghiện một thứ gì đó
20	rạn (san hô)	40	vũng nước

F. Word-meaning recall test

Vocabulary Test 1.1 (PRT)

Full Name:

Class:

Age:

Gender:

INSTRUCTION (Hướng dẫn):

Trong bài kiểm tra này, các em sẽ nghe **40 từ đơn lẻ**. Sau mỗi từ, các em sẽ có **20 giây** để **viết nghĩa Tiếng Việt** của từ đó. Nếu **chưa biết nghĩa** của từ đó, các em **đánh dấu X** bên cạnh số thứ tự tương ứng với từ đó. Dưới đây là một **ví dụ**:

Trong băng, các em nghe thấy: Number 1: Submarine

(Trong 20 giây)

Trong phiếu trả lời, các em viết:

1. tàu ngầm HOẶC 1. X

01		21	
02		22	
03		23	
04		24	
05		25	
06		26	
07		27	
08		28	
09		29	
10		30	
11		31	
12		32	
13		33	
14		34	
15		35	
16		36	
17		37	
18		38	
19		39	
20		40	

G. A sample of responses to the word-meaning recall test

Vocabulary Test 1.1 (IPT)			
		Gender: 	
<p>INSTRUCTION (Hướng dẫn):</p> <p>Trong bài kiểm tra này, các em sẽ nghe 40 từ đơn lẻ. Sau mỗi từ, các em sẽ có 20 giây để viết nghĩa Tiếng Việt của từ đó. Nếu chưa biết nghĩa của từ đó, các em đánh dấu X bên cạnh số thứ tự tương ứng với từ đó. Dưới đây là một ví dụ:</p> <div style="border: 1px solid black; padding: 10px; margin: 10px auto; width: 80%;"> <p>Trong băng, các em nghe thấy: Number 1: Submarine (Trong 20 giây) Trong phiếu trả lời, các em viết: 1. tàu ngầm HOẶC 1. X</p> </div>			
01	☒	X	
02	đường gây ngạc nhiên/sửng sốt.	✓	
03	rặng (sân hô).	✓	
04	X	X	
05	X	X	
06	thuộc địa	✓	
07	X	X	
08	rỗ / thoáng lỗ li vật gần gần cái	X	
09	X	X	
10	số chân (1 động vật có)	X	
11	di xương.	✓	
12	X	X	
13	X	X	
14	X	X	
15	X	X	
16	X	X	
17	thời miên / li thời miên.	✓	
18	gây giết mình.	✓	
19	đảm / chạc	X	
20	X	X	
21	X	X	
22	ánh sáng phát sáng của sơn vật đen	✓	
23	ma cà rồng.	✓	
24	X	X	
25	X	X	
26	lập lánh.	✓	
27	X	X	
28	nguy trang	✓	
29	X	X	
30	hàng điều khiển trong trò chơi điện tử	✓	
31	X	X	
32	X	X	
33	người say mê, yêu thích 1 thứ gì	✓	
34	X	X	
35	X	X	
36	hôn hợp / hôn lẩn	X	
37	thực hiện / hợp tác.	✓	
38	gây tức giận / giận dữ.	X	
39	X	X	
40	đảm bảo	✓	

H. Text comprehension test

The longest video

LISTENING COMPREHENSION QUESTIONS:

For the Third and Fourth Interventions – Repeated Viewing and Repeated Viewing plus Pushed Output

Your name:

Gender:

Class:

INSTRUCTION (Hướng dẫn):

For each of the statements below (*Với mỗi câu dưới đây*):

- Write YES if the information agrees with what is presented in the talk that you have watched
(Ghi Yes nếu nội dung câu đó trùng với nội dung trong video các em vừa xem).
- Write NO if the information is contradictory or irrelevant to what is presented in the talk
(Ghi No nếu nội dung câu đó trái ngược hoặc không được đề cập trong video các em vừa xem).

Section 1: Statements	Your answers
A. The presenter starts her talk with a fact that most sea animals produce light.	
B. She began to study the light-producing ability of sea animals when she was tracking pollution worldwide.	
C. The phenomenon that animals produce light has been observed more often in the ocean than any other places on earth.	
D. On the TED Conference stage, the presenter shows how an octopus makes light.	
E. One reason why sea animals produce light is to attack their enemies.	

Section 2: Statements	Your answers
F. Black Dragonfish can light up most of their organs as a call for help when in danger.	
G. Some shrimp can release glowing ink clouds to blind enemies.	
H. Fire Shooter can even fire to attack other sea animals.	
I. Golden bushes can change their colour from green to blue when we press them.	
J. The presenter suggests that some fishing technique should be banned to preserve precious sea creatures.	
K. There is evidence that living things on other planets can also produce light.	

Section 3: Statements	Your answers
L. An electronic jellyfish has been installed at the bottom of the sea to stimulate and capture the light-producing ability of sea animals.	
M. The presenter believes that it is this electronic jellyfish that has made some pearls release glowing chemicals.	
N. A Humboldt squid has even tried to rescue this electronic jellyfish.	
O. However, this electronic jellyfish is also found to be annoying to some sea animals.	

All three videos

LISTENING COMPREHENSION QUESTIONS

For the first and second interventions – Narrow Viewing and Narrow Viewing plus Interactional Output

Your name:

Gender:

Class:

INSTRUCTION (Hướng dẫn):

For each of the statements below (*Với mỗi câu dưới đây*):

- Write YES if the information agrees with what is presented in the talk that you have watched
(Ghi **Yes** nếu nội dung câu đó **trùng** với nội dung trong video các em vừa xem).
- Write NO if the information is contradictory or irrelevant to what is presented in the talk
(Ghi **No** nếu nội dung câu đó **trái ngược hoặc không được đề cập** trong video các em vừa xem).

Talk 1: Statements	Your answers
A. The beauty of the light that sea animals produce has been a source of inspiration for many art works.	
B. Most sea animals have an incredible eye sight.	
C. Some sea creatures can use their visual ability to sense the surrounding environment and blend themselves into it.	
D. Some can divide their body colour to attract mates and avoid enemies at the same time.	
E. Others can hide themselves behind moving rocks to avoid being seen.	

Talk 2: Statements	Your answers
F. Human beings have now explored more than 5% of the ocean.	
G. To explore the ocean, we need to have talent, collaboration, technical and financial support.	
H. In a dangerous situation, some squid can also produce light.	
I. In the deepest part of the ocean, where no light can reach, life still exists.	
J. Now we can stay at home and virtually explore the Titanic ship.	

Talk 3: Statements	Your answers
K. The presenter starts her talk with a fact that most sea animals produce light.	
L. She began to study the light-producing ability of sea animals when she was tracking pollution worldwide.	
M. According to the presenter, sea animals produce light mainly to protect themselves from enemies, attract mates, and find food.	
N. For example, some shrimp can release glowing ink clouds to blind their enemies.	
O. An electronic jellyfish has been installed at the bottom of the sea to stimulate and capture the light-producing ability of sea animals.	

I. A sample of note sheets

predators need to hide themselves.

Video 1: deep sea. *bioluminescent* (like firefly)
 pics of sea creature $\rightarrow$ blinking
 producing lights

avoid by eaten
 attract prey
 hypnotize &
 produce danger

wonder: deep sea is wonderful.
 creatures can change colour
 experiment: / males squid light up / male change colour $\rightarrow$ animal camouflage

Video 2: the deep & shallow deep sea

15 empire state building

we need to explore more about our sea $\rightarrow$ we don't know Earth
 account for 95% of Earth
 photos of sea: volcanoes / ponds / lake / river / in the sea (Mexico)

Atlantic Ocean: Titanic. sea creatures moved in.

[Researchers' dream help us know about Titanic from home.
 joystick

We explore 5% of ocean. 95% ? +

coral: amazing full of life $\rightarrow$ need to be protected

(deep) volcanoes: 80% deep in ocean

~~each time~~ submarine $\rightarrow$ ponds / lake / river

middle of the ocean is a mountain range
 largest waterfall in ocean. Iceland

submarine (3 pp) team	new species of life in ocean
2.5 $\rightarrow$ to bottom	hot spring: discover revolutionize
2h: pitch back	big animal
mid water: find life	vampire squid too many

J. Samples of summary and use of target words

RVS314: I think she is talking about the animals in the ocean, especially the animals that can produce light. She says that they use light to communicate. In two examples, she uses **optical lure**. When this **optical lure** started producing light, four or five fish came along and they started to produce light and together light up for communicating. And then a shark came. She said she didn't know it was because of this light or the **bait** that she put it in the **optical lure** that **lured** the shark. In the second example when she used the **optical lure**, a squid came, a squid or octopus I don't remember. First, it appeared to be shy. But later it came nearer and nearer to the **lure**. And I think that is how they communicate. He might think this **optical lure** was one of his kinds and he tried to talk to it, I think. And in the first part, she [the presenter] showed how these animals produced light in any part of their body. There were a lot of light colours like blue, green made by seahorse or some fish. I think that's it.

NVS532: Excuse me, NVS531. Can you be happy to tell me what you have seen in the video?

NVS531: I have seen two videos. The first one is about the intelligent design. The main point is about **camouflage**. This is the **camouflage** of animals living in the ocean. Uh. I don't know just some minor points. The first one is about the animals with incredible eyes.

NVS532: Can you elaborate more on that?

NVS531: Ah, I just hear that they....I don't know. So there are two parts about **camouflage**. The first one is, I don't what name it is, but that the animal settles in the surrounding environment and then disappear into the background. And the second one is about the two males. They fight with each other. It's about a male and female fish and then the male tries to hide his aggressive side from the female just by switching his side.

NVS532: What is the point of the video? Can you make more understandable?

NVS531: Yeah. It's about **camouflage**. But they just give minor examples.

NVS532: What is the point of **camouflage**?

NVS531: They say that this **camouflage** is amazing ...[unintelligible]. And in the end of the video, there is an octopus who change his pattern, his brightness, his colour and texture just to avoid the predators. So **camouflage** aims at avoiding the predators, enemies. That's all.

And the second video is about the ocean. The speaker said that the ocean is unexplored. We know just 5% of it. Meanwhile, water comprises 70% of our earth. And we can see hundreds of living animals from the bottom to the surface of the ocean. But to explore it, we need technology and a submarine is an example of technology. In a submarine, it takes 10 hours for people to look around...for the people in this submarine to travel in the ocean with none around. Then it will be very difficult. And they also discover that the life in the ocean is more diverse and denser than in a rainforest. That's all. I don't know. So do you have any questions?

NVS532: Can you tell me the point of the second video?

NVS531: This will help us to see the sea differently. It will help us to be aware of the values of the ocean and to discover it more.

NVS532: Ok.

NVS531: How about you?

NVS532: So I only watched one video. It's about **bioluminescence**, say, the glowing of living creatures. Specifically, the creatures under the depth of the ocean, about 2 or 3,000 feet under the ocean. In the video, she first showed some video about the glowing of animals under the depth of the ocean when she first dived, **plunged**, she first took a **plunge** and then she was amazed by it. She made a point that those glowing lights have...are used intentionally, just like a scream for help, to grasp attention or to **lure** a prey or predators or something like that. She showed some examples of shrimp, a squid, a fish and a lot of glowing animals. Oh when she caught a really scary black dragon fish, when they put it in the dark, all of its body glows, like on eyes, body, the tail. She says this's just like a scream for help because it is captured. And she did an experiment. She put **a glowing machine** or some signal machine under the ocean. And the animals surround this. They see it like a chatroom as they see it glowing, glowing just like they are sending each other a signal. And basically she concluded that it's attractive and addicted.

NVS531: Addicted?

NVS532: **Bioluminescence** and diving to the depth of the sea to see **bioluminescence** is like really addictive because it looks really beautiful.

K. Themes emerged from qualitative data

For the benefits of the summary activity in the RVS condition

Themes		
Layer 1	Layer 2	Layer 3
Benefits of the video	For text comprehension	<ul style="list-style-type: none"> - Textual clues - Pictorial clues - Audio clues
	For vocabulary learning	<ul style="list-style-type: none"> - Textual clues - Pictorial clues - Audio clues
Benefits of the summary activity	For text comprehension	Reprocess perceived information in order to: <ul style="list-style-type: none"> - decide what to include in the summary; - distinguish main ideas from supporting ideas; - organize the summary; - notice deficiencies in previous understanding and create purposes for the second viewing; - render processed information more memorable; - use the main ideas in the summary as signposts to facilitate the second viewing; - increase their confidence of text comprehension
	For vocabulary learning	<ul style="list-style-type: none"> - Notice the lexical gaps and create purposes for the second viewing; - Notice the lexical holes and create purposes for the second viewing; - Self-evaluate the attained knowledge of new words and decide not to experiment with using them, but paraphrase them instead; - Experiment with using the newly met words
Benefits of the second viewing	For text comprehension	<ul style="list-style-type: none"> - Confirm the reported information in the summary; - Elaborate the perceived information; - Correct the wrong information; - Add the information missed in the first viewing; - Improve the understanding of information organization
	For vocabulary learning	<ul style="list-style-type: none"> - Free working memory from focusing on text comprehension to focusing on processing novel words; - Direct their attention to the lexical items that they find they do not have through the summary activity; - Attempt to do form-meaning mappings for these words using: <ul style="list-style-type: none"> + the word formation; + the other learnt word meaning; + the co-textual clues; + the contextual clues

For the benefits of the summary activity in the NVS condition

Themes		
Layer 1	Layer 2	Layer 3
Benefits of the video	For text comprehension	<ul style="list-style-type: none"> - Textual clues - Pictorial clues - Audio clues
	For vocabulary learning	<ul style="list-style-type: none"> - Textual clues - Pictorial clues - Audio clues
Benefits of making a summary of the video content to a classmate	For text comprehension	Reprocess perceived information in order to: <ul style="list-style-type: none"> - decide what to include in the summary; - distinguish main ideas from supporting ideas; - organize the summary; - notice deficiencies in previous understanding; - render processed information more memorable; - increase their confidence of text comprehension
	For vocabulary learning	<ul style="list-style-type: none"> - Notice the lexical gaps; - Notice the lexical holes; - Self-evaluate the attained knowledge of new words and decide not to experiment with using them, but paraphrase them instead; - Experiment with using the newly met words
Benefits of listening a summary of the video content before actually viewing the video	For text comprehension	<ul style="list-style-type: none"> - Use the given topic to activate the top-down processing; - Establish an advance organizer of the upcoming information by using the key words or key events as the signposts to facilitate their actual viewing of the video; - Form some hypotheses of the upcoming information which facilitate their actual text comprehension as they merely need to: <ul style="list-style-type: none"> + confirm the hypotheses; + confirm the hypotheses and then elaborate them with further information; + reject the hypotheses and then form brand new hypotheses; + form brand new hypotheses that have not been mentioned yet
	For vocabulary learning	<ul style="list-style-type: none"> - Recycle learnt vocabulary knowledge; - Learn new words from their classmate
Benefits of the interaction in the summary activity	For text comprehension	<ul style="list-style-type: none"> - Seek for content clarification; - Elaborate ideas to get the intended message across
	For vocabulary learning	<ul style="list-style-type: none"> - Negotiate the form of new words; - Negotiate the meaning of new words