

# USING RETELLING AS A LISTENING-BASED OUTPUT TASK TO FOSTER L2 LISTENING COMPREHENSION AND INCIDENTAL VOCABULARY UPTAKE

Mr. NGUYEN Chi Duc, Ms. LUONG Quynh-Trang,  
& Mr. PHAM Xuan-Tho  
*ULIS, Vietnam National University*

## ABSTRACT

*This study examined the effect of an oral summary activity as a listening-based output task on L2 listening comprehension and incidental vocabulary uptake. In the same meaning-focused listening lesson of 45 minutes, EFL learners were required to watch a TED Talks video twice or watch this video twice, but with an oral summary activity inserted between the two viewing times. Their text comprehension was measured by 15 Yes/No questions. Their vocabulary gain was gauged by a word-form recall test and a word-meaning recall test. The combined effect of the repeated viewings and the summary activity indeed helped the learners achieve a significantly better text comprehension and greater incidental vocabulary uptake than the effect of the repeated viewings alone.*

**Keywords:** L2 listening comprehension, Incidental vocabulary uptake, Input repetition, Oral summary, TED Talks

## INTRODUCTION

An important added value of a meaning-focused L2 listening activity is its provision of opportunities for incidental L2 vocabulary learning (van Zeeland, 2013). Therefore, instructional practices that foster incidental vocabulary uptake in addition to fostering listening comprehension skills should be welcomed, perhaps especially if these practices require little extra investment of classroom time to harvest this added value.

However, L2 listening comprehension and incidental vocabulary learning from listening input are both challenging due to two crucial factors. One is the limited capacity of human working memory (Robinson, 2003; Schmidt, 1990). When L2 learners have used most of their working memory capacity for the purpose of text comprehension, their remaining mental resources are often inadequate for them to ponder unfamiliar words in the input text (Vandergrift, 2004; van Zeeland, 2013). The other, related, factor is the real-time nature of listening (Buck, 2001), where L2 learners do not have control over the speed at which the input text is presented to them (Perez, Noortgate, & Desmet, 2013) and consequently have insufficient time to invoke their explicit knowledge to supplement their online text processing (Hulstijn, 2007). As there is often no possibility of revisiting past passages, learners may not be able to verify their understanding of the content (Goh, 2000; Rost, 2013), let alone pay attention to how this content is worded (Boers, Eyckmans, & Godfroid, 2007; van Zeeland, 2013).

To help L2 learners overcome these two challenges and thereby facilitate their L2 listening comprehension and incidental vocabulary uptake in the same lesson, we adopt three suggestions in this study: (a) using audio-visual input (in the form of videotext) rather than audio-only input as study material, (b) exposing L2 learners to this material twice rather than once, and (c) inserting an output

activity between the two times of exposure to this material. It is the third suggestion the merits of which will be evaluated in the experimental study we report further below. We will first present the rationale for the first two suggestions as well, because they are implemented in the study, albeit not as independent variables.

There is ground to believe that using audio-visual rather than audio-only input as study material benefits both L2 listening comprehension and incidental vocabulary uptake. Previous research shows that audio-visual input brings about better L2 listening comprehension than audio-only input (e.g., Baltova, 1994; Shin, 1998; Wagner, 2010). There is also evidence that L2 learners are able to incidentally acquire new lexical knowledge from meaning-focused listening activities with audio-visual input materials (e.g., Neuman & Koskinen, 1992; Rodgers, 2013; Sydorenko, 2010; Vidal, 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 learners' vocabulary gain. The L2 learners in Sydorenko's (2010) study indeed reported using pictorial clues in the given audio-visual input material in their efforts to interpret unfamiliar words. Nevertheless, it needs to be acknowledged that the size of vocabulary gain from videotexts, especially ones with no or little modification, tends to be modest. In Rodgers' experiment (2013), for example, L2 learners in his experimental condition gained on average only six out of 60 target words after spending 13 weeks watching 13 episodes of an entertaining TV series (10%). Meanwhile, those in his control condition who had no exposure to these episodes in that period also gained on average five out of these 60 target words (8.33%). Similarly, Vidal (2011) found that her L2 learners only obtained a mean gain of 5.57 out of 36 target words (15.48%) after viewing three video-recorded lectures, and this declined to 7.81% one month later. Together, this review suggests that audio-visual input is more conducive to L2 listening comprehension and incidental vocabulary learning than audio input alone, but the size of vocabulary gain from this input still remains relatively small. Therefore, additional steps to enhance the rate of learning from audio-visual materials would be welcome.

The second aforementioned suggestion, i.e., of listening to the same input text twice rather than once, addresses the second problem that L2 learners face – the real-time nature of listening. When learners are given the opportunity to revisit the same listening passage, they can verify their previous interpretation of the content and add information they did not capture in the first listening. Repeated listening has long been considered a useful technique to make listening input more comprehensible (Krashen, 1996; Vandergrift & Goh, 2012, pp.156-157), and research findings support this view (e.g., Chang & Read, 2006; Elkhafaifi, 2005; Saika, 2009). Given the enhanced text comprehension, it is expected that L2 learners will have more mental resources to attend to how the content is worded (Boers et al., 2007; van Zeeland, 2013) and more contextual clues for their interpretation of unfamiliar words (Stæhr, 2009). Moreover, listening to the same text twice entails learners' re-encountering the same words. Frequency of word encounters is one of the predictors of incidental L2 vocabulary acquisition from reading and listening (Brown, Waring, & Donkaewbua, 2008; van Zeeland & Schmitt, 2013; Vidal, 2011). 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 inferences thanks to an accumulation of contextual clues. Still, given the above discussion, it remains reasonable to expect a better listening comprehension level and incidental vocabulary uptake from listening to the same audio-visual input twice than just once. At the same time, we do recognize that learners ultimately need to be equipped with the skills and strategies that enable them to cope with 'real-life' listening conditions where re-listening to the same input is often not an option (Vandergrift & Goh, 2012, pp. 166-197).

Let's now turn to the third aforementioned suggestion, i.e., that of inserting an output activity between the two times of exposure to the same input text. As explained further below, this will constitute the independent variable in the experimental study reported in the present article. The output activity that is the object of this study is an oral summary task, i.e., where learners sum up the content of the listening text before they listen to the text a second time. We hypothesize that this offers advantages for L2 listening comprehension and incidental L2 vocabulary learning over and above the two favourable

conditions already discussed (i.e., using audio-visual input and repeated listening). Regarding L2 listening comprehension, the announced task of summing up the content may prompt more engagement with the input text (Stahl & Clark, 1987). In addition, while summarizing the content after they have listened to the material once, learners may notice deficiencies in their understanding of the input, which, in turn, may prompt them to attend to relevant information during their second listening. The summary task can also stimulate vocabulary learning in accordance with two core tenets of Swain's Output Hypothesis. First, this activity promotes "noticing" (Swain, 1995, p.125). When L2 learners are required to summarize the input text, they may notice gaps in their lexical resources and subsequently direct their attention to relevant lexis as they are exposed to the same text again. The second tenet, which has perhaps received less attention in the research literature, is that output enhances fluency of L2 use. The activity provides learners with opportunities to experiment with using newly met words, which may help create more durable memories of them. In short, there are grounds for expecting the proposed summary activity to be beneficial for learners' engagement with text content as well as the (lexical) packaging of that content. Empirical evidence in support of this expectation has been furnished in connection with reading-based output activities (e.g., Joe, 1998; Rassaei, 2015), but, to our knowledge, not yet in connection with listening input.

Below, we report a classroom-based experiment with a mixed within- and between-participant research design intended to compare the degree of L2 listening comprehension and incidental vocabulary gain in two treatment conditions: (a) exposure to audio-visual material twice, (b) exposure to this input twice, but with an oral summary activity incorporated in between. These conditions will henceforth be referred to respectively as the 'repeated viewing' and the 'repeated viewing plus summary' condition, or, for short, as the RV and the RVS condition. In both conditions, the first two suggestions we discussed above (i.e., using audio-visual input and repeated exposure) were implemented as common features. The independent variable under examination is the summary activity which distinguishes the RVS from the RV condition.

To better evaluate the merits of the treatments regarding vocabulary gains, we also included a control condition where participants took the vocabulary pre- and post-tests, but did not get to listen to the audio-visual material. This will enable us to gauge the effects of mere test-taking and of chance learning of target words outside of the experiment.

The practical relevance of the experiment is that its outcome may help inform practitioners' design of lessons that include listening input. More specifically, if it emerges from the experiment that the RVS condition generates significantly better learning outcomes, then this finding should encourage practitioners to incorporate output activities (such as oral summary activities) in their listening lessons, especially if the lessons' objectives include the development of listening comprehension skills *and* incidental vocabulary uptake.

### ***Research Questions***

The study seeks answers to the following primary research questions:

- (a) Is better comprehension of the audio-visual text obtained in the RVS condition than in the RV condition?*
- (b) Is more vocabulary knowledge gained in the RVS condition than in the RV condition?*

Additionally, the test data will be examined for the strength of association between vocabulary acquisition, on the one hand, and (i) listening comprehension and (ii) frequency of word occurrences in the listening input, on the other. Such associations have been reported before (e.g., Sawada, 2009; Vidal, 2011), but have not yet been explored in the treatment conditions examined in the present study.

## MATERIAL AND METHOD

### *Research participants*

The participants were 97 Vietnamese learners (three males and 74 females) of English as a foreign language from a university in Vietnam. 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 CEFR C1 level or IELTS overall band score of 6.5. These participants were, in their intact classes, assigned to the RV ( $n = 33$ ), the RVS ( $n = 32$ ), and the control ( $n = 32$ ) condition. Prior to this experiment, their English language proficiency, as measured by a series of in-house English language proficiency tests using CEFR as the underlying test construct, was CEFR B2+ or IELTS overall band score of 5.5 (i.e., upper intermediate). Scores on Nation and Beglar's (2007) Vocabulary Size Test (VST) indicated that these learners had receptive knowledge that included the 4,000 most frequent word families in English. Because not all test data were normally distributed, we resorted to non-parametric statistics to ascertain that there were no significant differences between the groups. Kruskal-Wallis tests (i.e., the non-parametric counterpart of ANOVA tests for independent samples) indeed produced  $H(2,97) = 0.71$ , ( $p = .700$ ) and  $H(2,97) = 1.93$  ( $p = .382$ ) for the overall scores on the proficiency tests and the scores on the VST, respectively.

### *Study material*

TED Talks were chosen as the audio-visual input material in this experiment. These talks (typically ranging from four to twenty minutes long) are lively and freely available web-based presentations on a wide variety of professional and academic topics, delivered by experts who attempt to make these topics accessible for a general audience. Given the pedagogy-oriented nature of the study, we opted for authentic audio-visual materials that are both readily available and manageable for classroom use in terms of length. TED Talks appear to meet these requirements. In addition, previous research has shown that TED Talks have good potential for both L2 listening and vocabulary development (Coxhead & Walls, 2012; Takaesu, 2013). Finally, a growing number of language institutions worldwide have started using these presentations in their programmes (Coxhead & Walls, 2012). Also the policy makers at the institution where we collected our data were considering incorporating TED Talks in the curriculum at the time our study was conducted. This made our choice to use TED Talks in our experiment both timely and ecologically justified. We opted for TED Talks without captions, because our study examined L2 listening comprehension and incidental vocabulary uptake from a listening procedure, not a reading procedure.

Sixteen learners with a similar L2 profile to the participants recruited for the actual experiment in this study 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, a talk was chosen. This talk was about marine bioluminescence and lasted twelve minutes. From the RANGE program outcome, 20 words beyond the most frequent 4,000 word families in English were selected as the potential target items. Choosing these words reduced the likelihood that the participants might already know these items prior to the experiment. These words varied in frequency of in-text occurrences (from one to six times), in part of speech (e.g., nouns, verbs, and adjectives) and in length (from one (e.g., *plume* and *bait*) to four syllables (e.g., *contemplative* and *unobtrusive*)). The target words, their number of in-text occurrences, and a link to this video are provided in Appendix A.

### *Procedure*

The participants in the RV condition were required to watch and take notes of the content of the selected video twice (25 minutes) with a view to answering content questions afterwards. After each viewing time, they had five minutes to make modifications and additions to their notes. Finally, they were given 15 content questions and were allowed to consult their notes while answering these questions (10 minutes). Those in the RVS condition followed exactly the same procedure as those in the RV condition, but, instead of expanding and/or elaborating their notes, they used the five minutes after their first viewing to summarize the video content. They were asked to audio-record this summary. The learners in

the control condition followed the same procedure as those in the RV condition, but with a different TED Talks video which included none of the target words. The time invested was kept the same, at 45 minutes, across all conditions. The participants were informed beforehand that content questions would follow, but the vocabulary test was not announced (since the study is an examination of incidental, not deliberate, vocabulary learning).

### ***Text comprehension measure***

According to Lynch (2011), presentations such as TED Talks belong to the realm of one-way academic listening, as they bear resemblance to university lectures and conference presentations. To gauge learners' comprehension of this kind of material, guidelines proposed in the literature on L2 academic listening are therefore useful. Wagner (2002) describes the construct of L2 academic listening with reference to the well-established two-pronged model of top-down and bottom-up processing (Buck, 2001, pp. 02-03; Flowerdew, 1994, pp. 09-13; Nation & Newton, 2009, pp. 40-42). After revisiting the tasks used in earlier work on academic listening (Buck & Tatsuoka, 1998; Weir, 1993), he argues that the most valid way of measuring L2 academic listening comprehension is a combination of tasks that require learners to: (a) identify the purpose and main ideas of the text, (b) make inferences about implicit meaning, logical links between ideas, and speaker attitude, and (c) recognize important supporting ideas and factual information. Thus, we designed our text comprehension test with these components in mind.

Fifteen content questions in a YES/NO format were used to gauge the extent to which the participants understood and remembered the video content. Three experienced EFL/ESL teachers (one a native speaker and the others non-native speakers of English) were invited to watch the selected video and then read its transcript to determine these three types of information. Fifteen sets of information proposed by all three teachers were then chosen to design the test. Following Field (2012), these questions were all presented in the form of statements for the learners to decide whether or not they reflected the video content, by writing "Yes" or "No" next to each statement. Below are three examples of these statements: "*One reason why sea animals produce light is to attack their enemies*" (main idea), "*There is evidence that living things on other planets can also produce light*" (implicit meaning), and "*A Humboldt squid has even tried to rescue this electronic jellyfish*" (factual information). In the scoring procedure, 1 and 0 points were awarded to correct and incorrect answers, respectively.

### ***Vocabulary measures***

The vocabulary knowledge measures in this study were two different tests. One was in a word-meaning recall format and the other was in a (spoken) word-form recall format. To prepare the former, a male native speaker of English was asked to read aloud and to audio-record the 20 target words with an interval of 20 seconds between the words. The participants were 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 were presented in Vietnamese (L1) on the response sheet. To prepare the word-form recall test, the aforementioned male native speaker of English read aloud and audio-recorded the initial syllables of the target words with an interval of 20 seconds between them. The participants were asked to say the target word aloud and audio-record it, prompted by (a) the target word's meaning written in Vietnamese on the test sheet and (b) the first syllable of the target word 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*") were missing from the prompts. An experienced Vietnamese teacher of English checked the accuracy of L1 translations of the target words. During the word-form recall test, the participants wore an *over-ear* headphone with a microphone attached. This device helped them listen to the test prompt and audio-record their response without being distracted. Only four participants took the test at a time and they were seated at different corners of a large classroom. It took the participants 15 minutes to complete each of the tests.

Both receptive and productive knowledge were tested in this study since there is a consensus in the field that vocabulary knowledge includes more than one dimension (van Zeeland & Schmitt, 2013) and that some dimensions (e.g., word form) might be more readily learnt than others (e.g., word meaning) (van Zeeland & Schmitt, 2013; Brown et al., 2008). Word-meaning recall tests have already been used in such

studies as van Zeeland and Schmitt (2013) and Brown et al. (2008). However, in those studies the test prompts were presented in the written form even though not all the participants had been exposed to the written form of the target words in the input materials. Sydorenko (2010) found a significant difference in the vocabulary test scores when the presentation mode of the test prompt was either congruent or not congruent with that of the input text. Therefore, the word-meaning recall test in the current study used the spoken rather than written word form as the prompt for word recall. Tests on word form have also been used in previous research (e.g., Sydorenko, 2010; van Zeeland & Schmitt, 2013). However, these focused on the participants' ability to recognize rather than recall the word forms. In this study, we examined whether the participants were able to recall the forms of lexical items they had heard in the audio-visual input. The use of the L2 first syllables was to confine responses to the target words rather than possible synonyms.

Had the participants been given both of these vocabulary tests, their exposure to the test prompt in one test might have primed their response in the other. Specifically, if they had taken the word-form recall test first, they might have familiarized themselves with the target-word meaning before taking the word-meaning recall test. Conversely, if they had been provided with the word-meaning recall test first, exposure to the L2 word-form prompt would have helped 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 had been assigned to), the word-meaning recall test in this study was assigned to half of the participants in each treatment condition while the word-form recall test was given to the other half. As a result, for the vocabulary measures, there were two different datasets (one for the word-meaning recall test and the other for the word-form recall test).

The 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 “*môi 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 that by the primary investigator of this study. No differences were found in any between-rater comparisons of the awarded scores.

### ***Data analysis***

The purpose of our vocabulary pre-test was to establish which among our pre-selected potential target words were unknown to the learners prior to the experiment. Their pre-test scores showed that three words (i.e., “*breath-taking*”, “*plunge*”, and “*twinkling*”) on the word-form recall test and two words (i.e., “*breath-taking*” and “*twinkling*”) on the word-meaning recall test were already known by 10% or more of the learners. Following the studies by Ellis and his colleagues (1999), these target words were omitted from the final analyses. After this exclusion, out of the original potential 20 target words, there were now 17 actual target words in the word-form recall test and 18 actual target words in the meaning-recall test.

We shall use the term “*acquisition*” to refer to vocabulary learning as evidenced in the immediate post-test, and the term “*retention*” to refer to vocabulary learning as evidenced in the delayed post-test. 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. Because not all test data were normally distributed, we shall resort to non-parametric inferential statistics to compare the effects of the two learning conditions on L2 listening comprehension incidental vocabulary uptake.

## RESULTS

### *Text comprehension level: RVS > RV?*

Table 1 presents the descriptive statistics including the sample size ( $N$ ), the mean score ( $M$ ), and the standard deviation ( $SD$ ) for the text comprehension scores per condition:

Table 1: Text comprehension scores

	$N$	$M$	$SD$
RV	33	7.15	2.28
RVS	32	8.94	2.26

According to the Mann Whitney test, the RVS group outperformed the RV group:  $Z = 2.96$  ( $p = .003$ ).

### *Incidental vocabulary gain: RVS > RV?*

As already mentioned, we included the control condition to gauge the effects of test-taking and chance learning of the target words outside of the experiment. Therefore, it is worth examining these effects first. Table 2 presents the descriptive statistics for the vocabulary test scores of the control condition per vocabulary measure, including the mean scores outside brackets and the corresponding standard deviations inside brackets:

Table 2: Vocabulary gain in the control condition

	Form recall (max. = 17 points)	Meaning recall (max. = 18 points)
Pre-test	0.07 (0.27)	0.33 (0.49)
Immediate post-test	0.07 (0.27)	0.33 (0.49)
Delayed post-test	0.07 (0.27)	0.33 (0.49)
Acquisition	0.00 (0.00)	0.00 (0.00)
Retention	0.00 (0.00)	0.00 (0.00)

It is clear from Table 2 that there were no discernible effects of test-taking and chance learning of the target words in this experiment were null. Therefore, it is now safe to attribute vocabulary gains (if any) in the treatment groups to the suggested learning conditions.

Table 3 presents the descriptive statistics for the vocabulary scores on the word-form recall measure per condition:

Table 3: Vocabulary scores on word-form recall measure (max. = 17 points)

	RV ( $n = 14$ )	RVS ( $n = 14$ )
Pre-test	0.29 (0.47)	0.31 (0.60)
Immediate post-test	1.14 (0.95)	2.25 (2.05)
Delayed post-test	1.07 (0.92)	2.31 (2.02)
Acquisition	0.85 (0.77)	1.94 (1.64)
Retention	0.78 (0.70)	2.00 (1.62)

The result from a Mann Whitney test showed that there was no significant difference in the pre-test scores between these groups:  $Z = 0.21$  ( $p = .834$ ). Three Friedman tests were then run to investigate the difference in the vocabulary test scores per condition across the three testing times – the pre-test, the immediate post-test and the delayed post-test. In the RV condition, a significant difference was found:  $X^2(2,14) = 8.71$  ( $p = .013$ ). Three Wilcoxon Signed-Rank tests were run for the pairwise comparisons and indicated that the RV group's scores in the two post-tests were both significantly higher than the pre-test scores even when the post-hoc Bonferroni correction was applied, with the same  $Z = 2.67$  ( $p = .004 < .017$ ). A significant difference was also found in the case of the RVS condition:  $X^2(2,14) = 18.41$  ( $p < .0001$ ). The results from the Wilcoxon Signed-Rank tests also revealed here that both post-tests significantly surpassed the pre-test scores with the same  $Z = 3.30$  ( $p = .0001$ ). This indicates that both treatment conditions led to a significant vocabulary gain. To investigate the difference in the size of the vocabulary gain between these conditions, two Mann Whitney tests were run for their vocabulary acquisition and retention scores. Both vocabulary acquisition and retention scores in the RVS condition were found to outweigh those in the RV condition, with  $Z = 2.00$  ( $p = .046$ ) and  $Z = 2.37$  ( $p = .018$ ), respectively.

Table 4 presents the descriptive statistics for the vocabulary scores on the word-meaning recall measure:

Table 4: Vocabulary scores on word-meaning recall measure (max. = 18 points)

	RV ( $n = 19$ )	RVS ( $n = 18$ )
Pre-test	0.26 (0.56)	0.17 (0.38)
Immediate post-test	1.00 (1.05)	1.94 (1.39)
Delayed post-test	0.84 (1.01)	1.89 (1.23)
Acquisition	0.74 (0.65)	1.77 (1.22)

---

Retention	0.58 (0.77)	1.72 (1.07)
-----------	-------------	-------------

---

The result from a Kruskal Wallis test indicated that there was no significant difference in the pre-test scores between the conditions:  $Z = 0.26$  ( $p = .795$ ). Three Friedman tests were run to investigate the difference in the vocabulary scores per condition across the three testing times. In the RV condition, a significant difference was found:  $X^2(2,19) = 8.55$  ( $p = .014$ ). Three Wilcoxon Signed-Rank tests were then applied for the pairwise comparisons and these indicated that the scores in both post-tests were significantly higher than the pre-test scores, and significance was reached even when the post-hoc Bonferroni correction was applied, with  $Z = 3.06$  ( $p = .002 < .017$ ) and  $Z = 2.52$  ( $p = .006 < .017$ ), respectively. Another significant difference was found in the case of the RVS condition:  $X^2(2,18) = 24.11$  ( $p < .0001$ ). The results from the Wilcoxon Signed-Rank tests revealed also here that the scores on the two post-test significantly surpassed the pre-test scores, with the same  $Z = 3.62$  ( $p = .001 < .017$ ). Again, this suggests that the two treatment conditions resulted in significant vocabulary uptake. To examine the difference in the size of the vocabulary uptake between the conditions, two Kruskal Wallis tests were run on the vocabulary acquisition and retention scores. Both vocabulary acquisition and retention scores in the RVS condition were found to be significantly higher than those in the RV condition, with  $Z = 3.09$  ( $p = .002 < .05$ ) and  $Z = 3.40$  ( $p = .0007 < .05$ ), respectively.

### *Frequency effect?*

Recall that a secondary research question concerned the predicted positive effect of multiple encounters with the same word. To ascertain whether this association also holds true in the present learning conditions, where the same words are re-encountered mostly in identical contexts, we computed Pearson correlation coefficients between word frequency and successful word recall rates. The results confirm the expected positive association. On the word-form recall measure, the frequency of word occurrences in the input text was positively associated with vocabulary acquisition ( $r = .61$ ,  $p < .01$ ) and the vocabulary retention scores ( $r = .58$ ,  $p < .01$ ). This is also the case for the word-meaning recall test, with  $r = .46$  ( $p < .05$ ) and  $r = .57$  ( $p < .01$ ), respectively.

## **DISCUSSION**

The above results corroborate the hypothesis that incorporating an output activity such as an oral summary activity between two exposures to the same listening (or audio-visual) input is beneficial for vocabulary uptake. In the introduction, several reasons were proposed in support of this hypothesis, including the likely positive association between engagement with text content and vocabulary uptake and also the possibility that learners might attempt to recycle newly encountered words in their own output. In what follows, we explore whether these are indeed plausible explanations for the quantitative findings.

### *Is vocabulary gain associated with text comprehension?*

To examine the association between the vocabulary gain on one hand and text comprehension on the other, we combined the test data from the two groups and computed Pearson correlation coefficients. On the word-form recall measure, the text comprehension scores were found to be positively correlated with the vocabulary acquisition scores ( $r = .54$ ,  $p = .0002$ ) and the vocabulary retention scores ( $r = .57$ ,  $p < .0001$ ). This positive association is paralleled for the word-meaning recall measure:  $r = .47$  ( $p = .0002$ ) and  $r = .49$  ( $p < .0001$ ), respectively. As the RVS condition produced the better comprehension test scores as well as the better vocabulary test scores (see above), these results are at least consistent with the proposition that the summary activity engaged learners with the content of the video in ways that benefited vocabulary uptake.

### ***Target word use in summaries?***

In this study, the incorporation of a summary activity between the two viewings of the same video was found to benefit vocabulary uptake. This is in accord with Swain's Output Hypothesis (1995). To examine whether the learners in the RVS condition might have enhanced their vocabulary gain thanks to their use of the target words in their summary, the summaries were transcribed and screened for the presence of target words. Altogether we found four out of the 18 target words in our collection of the summaries (*lure* (x9 tokens), *bait* (x1), *bioluminescence* (x2), and *saber* (x1)). Eleven of the 32 learners in the RVS group used one or more target words in their summaries. The total number of times these words occurred in the corpus of summaries was 14 (one learner used one target word three times and another target word once, while the other ten learners used only target word and did so only once). The words were not yet known prior to the treatment, according to the pre-test data, and yet they were used in the summaries in grammatically and semantically appropriate ways. Of direct interest for the question of whether the summary output task benefits vocabulary learning is the finding that the participants who used these target words in their summaries also recalled them in the two post-tests.

It may be of interest that the target words were not used as part of utterances that were recycled verbatim from the input text. The syntagmatic contexts in which the target words were used were all different from the original phrasings in the video. Below is one example:

Original context: "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."

S20: "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."

In the summaries, we 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, a hesitation which, is a likely sign that learners were looking for the target words (Tavakoli, 2011), but then failed to retrieve them from memory. After the pause, the participant 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 – again consistent with the tenets of the Pushed Output Hypothesis.

## **CONCLUSION, LIMITATIONS, AND IMPLICATIONS**

Regarding text comprehension, the learners in the RV condition and the RVS condition obtained a mean score of, respectively, 7.15 (48%) and 8.94 (60%) out of the maximum 15 points on the text comprehension test. One plausible explanation for the better score under the RVS condition is that the summary task required the learners to reprocess the information they had written down in their notes in order to produce a coherent and sufficiently informative summary. In this way, they may have improved their retention of content (Baddeley, 1990; Joe, 1998). In addition, the activity may have prompted them to notice gaps in their initial interpretation of the content, which, in turn, may have directed their attention to this information when they watched the talk a second time.

As to vocabulary uptake, the average pre-to post-test gains were about 1 (5.88%) and 2 (11.77%) target words in the RV and RVS conditions, respectively. It is clear that these gains are very modest. This limited vocabulary gain is similar to what such well-designed studies as Rodgers (2013) and Vidal (2011) have already reported. This also reflects the slow pace of incidental vocabulary learning in general (Nagy, Anderson, & Herman, 1985), and confirms the challenges of incidental L2 vocabulary learning

from listening input in particular (Vandergrift, 2004; van Zeeland, 2013), including – so it appears – TED Talks.

In the case of the present study, additional factors may help to account for the very modest gain. First, the study material in this experiment was authentic (in the sense that it was produced without L2 learners in mind (Coxhead & Walls, 2012)). Such characteristics of authentic materials as natural and fast speech delivery must be challenging to learners (Vandergrift, 2004). It is worth noting in this regard that most previous research on incidental vocabulary learning used either special-purpose instructional materials, like graded readers (e.g., Brown et al., 2008; Webb & Chang, 2012) or materials developed specifically for research purposes (e.g., van Zeeland & Schmitt, 2013; Vidal, 2011). For example, as the video in this experiment was not modified, word repetition in the input text was not manipulated either. Consequently, all but one target word (“*bioluminescence*”) occurred only once or twice in the text, which is much lower than the 10+ or 20+ times suggested by van Zeeland and Schmitt (2013, p. 621) or Brown et al. (2008, p. 153) as necessary for incidental vocabulary learning. Second, although our piloted learners suggested that the selected TED Talks video was suitable for use in this study (see section *Study material* above), the comprehension test scores indicate that its content was far from straightforward. When text passages are not well understood, unfamiliar words in those passages do not stand a good chance of being adequately interpreted either. Finally, TED Talks often contain technical words, denoting concepts that may be unfamiliar to the participants (Coxhead & Walls, 2012). The TED Talks video used in the present study was no exception (e.g., “*trawling*”, “*lure*” or “*barbell*”). Therefore, the learners may have needed to establish these new concepts in mind first rather than directly map known concepts to L2 forms (Vidal, 2011). It is also possible that the learners did not perceive the topic-specific target words as useful additions to their L2 resources.

It needs to be mentioned, however, that gains attested in the vocabulary tests we used are likely to underestimate the amount of acquisition that actually took place. For one thing, the learners in the experiment 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. The fact that our vocabulary test could not pick up signs of acquisition beyond the selected target words is one of the limitations to this study.

In any case, it needs to be borne in mind that incidental vocabulary uptake in this study is to be interpreted as added value to meaning-focused L2 listening practice, and so even modest uptake should be considered in a positive light. Moreover, the summary task added output practice to the lesson, thus creating more diverse skills development opportunities the potential benefits of which are beyond the scope of what we can discuss in this article.

Returning to the greater vocabulary gain under the RVS condition, this indeed appears a positive side-effect of the summary activity. The conspicuously long pauses signalling likely word-retrieval efforts (and failures) often followed by successful word recall in the post-tests suggests that learners noticed gaps in their lexical resources and filled these when they processed the input text a second time. When learners did manage to integrate newly encountered words in their own output, this was also followed by successful recall in the post-tests, which suggests another beneficial effect of the output task. As expected, the vocabulary uptake in this experiment was also found to be positively correlated with text comprehension and the number of word occurrences in the text, which echoes the findings of previous studies (e.g., Sawada, 2009; Vidal, 2011).

In conclusion, inserting an oral summary activity between the two times of viewing the same TED Talks video was found to foster better text comprehension and greater incidental vocabulary gain than viewing the video twice without an output activity. Specifically, in the same classroom time of 45 minutes, this incorporation of summary work helped L2 learners increase their text comprehension scores by 12% and their incidental vocabulary gain scores by 6.40%. Therefore, if L2 teachers would like to make better use of their classroom time in a meaning-focused listening lesson, asking their students to make an oral summary of the input before their re-exposure to this input is likely to be more fruitful than asking them to further elaborate their notes of the input content.

Apart from the aforementioned limitations to the vocabulary uptake measures, other limitations of this study need to be acknowledged. One is the small sample sizes, in terms of participants as well as test items. Another is that only one type of output task was investigated. While our findings support the Output Hypothesis, it is worth examining whether similar benefits may be reaped if other text-based output tasks are implemented (see, e.g., Rassaei (2015) in the context of reading), such as ones that involve learner-learner interaction instead of the monologic summary task we have tried here. Thus, one strand of research that could further inform procedures in L2 listening practice is to investigate and compare the benefits for text comprehension and vocabulary uptake of such alternative text-based output activities.

## REFERENCES

- Baltova, I. (1994) The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review* 50: 507-31.
- Boers, F., J. Eyckmans and A. Godfroid (2007) Catering for limited processing capacity to foster incidental vocabulary uptake. In K. Pelsmaekers and C. Rollo (eds.), *Economically Speaking: Essays in Honour of Chris Braecke*. Antwerp: Garant Publishers. 169-187.
- Brown, R., R. Waring and S. Donkaewbua (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.
- Buck, G. and K. Tatsuoka (1998) Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing* 15: 119-157.
- Caliendo, G. and A. Compagnone (2014) Expressing epistemic stance in university lectures and TED Talks: A contrastive corpus-based analysis. *Lingue e Linguaggi* 11: 105-122.
- Chang, A. and J. Read (2006) The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly* 40.2: 375-397.
- Coxhead, A. and R. Walls (2012) TED Talks, vocabulary, and listening for EAP. *The TESOLANDZ Journal* 20.1: 55-58.
- Elkhafaifi, H. (2005) The effect of pre-listening activities on listening comprehension in Arabic learners. *Foreign Language Annals* 38.4: 505-513.
- Ellis, R. (1999) *Learning a Second Language through Interaction*. Amsterdam: John Benjamin Publishing Co.
- 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.
- Goh, C. (2000) A cognitive perspective on language learners' listening comprehension problems. *System* 28.1: 55-75.
- Hulstijn, J. H. (2007) Psycholinguistic perspectives on language and its acquisition. In J. Cummins and C. Davidson (eds.), *International Handbook of English Language Teaching*. Norwell: Springer. 783-795.
- Joe, A. (1998) What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied linguistics* 19.3: 357-377.
- Krashen, S. (1996) The case for narrow listening. *System* 24: 97-100.

- Lynch, T. (2011) Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes* 10: 79-88.
- Nagy, W. E., P. A. Herman and R. C. Anderson (1985). Learning words from context. *Reading Research Quarterly* 20.2: 233-253.
- Nation, P. and D. Beglar (2007) A vocabulary size test. *The Language Teacher* 31.7: 9-13.
- Nation, P. and A. Heatley (2002) *Range: A program for the analysis of vocabulary in texts*. Available at <http://www.victoria.ac.nz/lals/about/staff/paul-nation> (accessed 10 June 2015)
- Nation, P. and J. Newton (2009) *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.
- Neuman, B. and P. Koskinen (1992) Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly* 27.1: 96-106.
- Perez, M., W. Noortgate and P. Desmet (2013) Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System* 41: 720-739.
- Rassaei, E. (2015) Effects of three forms of reading-based output activity on L2 vocabulary learning. *Language Teaching Research*. doi: 1362168815606160.
- Robinson, P. (2003) Attention and memory during SLA. In C. J. Doughty and M. H. Long (eds.), *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing. 631-678.
- Rodgers, M. (2013) *English Language Learning through Viewing Television: An Investigation of Comprehension, Incidental Vocabulary Acquisition, Lexical Coverage, Attitudes, and Captions*. Unpublished PhD thesis, Victoria University of Wellington, New Zealand.
- Rost, M. (ed.) (2013) *Teaching and Researching Listening* (2<sup>nd</sup> edition). Harlow: Pearson
- 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 PhD Thesis, Temple University, USA.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11.2: 129-158.
- Shin, D. (1998) Using videotaped lectures for testing academic listening proficiency. *International Journal of Listening* 12: 57-80.
- Stæhr, L. (2009) Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition* 31.4: 577-607.
- Stahl, S. and H. Clark (1987) The effects of participatory expectations in classroom discussion on the learning of science vocabulary. *American Educational Research Journal* 24.4: 541-555.
- Swain, M. (1995) Three functions of output in second language learning. In. G. Cook and G. Seidhofer (eds.), *Principles and Practices in Applied Linguistics: Studies in Honour of H. G. Widdowson*. Oxford, UK: Oxford University Press. 125-144.
- 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* 4.2: 150-162.
- Tavakoli, P. (2011) Pausing patterns: differences between L2 learners and native speakers. *ELT Journal* 65.1: 71-79.
- van Zeeland, H. (2013) Vocabulary and listening. *The Encyclopaedia of Applied Linguistics*. Australia: Wiley-Blackwell.
- van Zeeland, H. and N. Schmitt (2012) Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics* 34.4: 457-479.
- van Zeeland, H. and N. Schmitt (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. and C. C. Goh (2012) *Teaching and Learning Second Language Listening: Metacognition in Action*. New York: Routledge.
- 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. and A. C. Chang (2012) Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review* 68.3: 267-290.
- Weir, C. J. (1993) *Understanding and Developing Language Tests*. New York: Prentice-Hall.

## APPENDIX

Link to the selected video	Target words and frequency of their in-text occurrences
<a href="https://www.ted.com/talks/edith_widder_the_weird_and_wonderful_world_of_bioluminescence">https://www.ted.com/talks/edith_widder_the_weird_and_wonderful_world_of_bioluminescence</a>	<i>Propagate (1), plunge (2), mesh (2), bait (2), breathtaking (2), morsel (1), trawling (1), plankton (2), twinkling (1), contemplative (1), plume (2), barbel (2), barrage (2), unobtrusive (1), saber (1), entranced (1), lure (1), foolhardy (1), junky (1), bioluminescence (6)</i>

## AUTHORS

Ms. NGUYEN Chi Duc (BA honour in TEFL, Vietnam National University, Vietnam; MA in TESOL, Victoria University of Wellington, New Zealand; PhD in Applied Linguistics, Victoria University of Wellington, New Zealand) has been a teacher of English as a foreign language in Vietnam for over 10 years. His research interests include second language vocabulary acquisition and learning, academic listening, language testing and assessment. His latest publications will be

found in the TESOL Encyclopaedia of Teaching English Language edited by J. I. Lontas by the end of this year. Email: [Chiduc.nguyen@vuw.ac.nz](mailto:Chiduc.nguyen@vuw.ac.nz)

PHAM Xuan-Tho (*BA in TEFL, Vietnam National University, Hanoi; MA in Applied Linguistics, La Trobe University, Australia*) has been a lecturer in English Linguistics at University of Languages and International Studies, Vietnam National University, Hanoi for more than 20 years. His on-going research interests have been in the area of second language acquisition, inter-language phonetics, syntax and pragmatics.

LUONG Quynh-Trang earned her M.A. in TESOL in 2004 and M.A. in Education in 2010 from The University of Melbourne, Australia. She has 17 years of experience teaching English as a second language and ELT Methodology. Currently, she is Head of the ELT Methodology Division at the Faculty of English Language Teacher Education, ULIS, VNU, Hanoi. She has been an author and chief author of the new series of English textbooks for lower-secondary level in Viet Nam. Her interests include ELT methodology, teacher education and teacher professional development.