The Relationship Between Incidental Vocabulary Learning and Extensive Listening to Podcasts

Nare K. Hakobyan
American University of Armenia, Armenia

Abstract
The increasing popularity of listening to podcasts for the purpose of improving English among Armenian teenagers raises the question as to what impact this might have on incidental vocabulary learning. A positive impact would be very welcome, especially in Armenia, where podcasts might prove to be one of the few affordable technology applications. This prospect gave rise to a mixed-method, quasi-experimental study designed to answer the following research questions: (1) What is the relationship between incidental vocabulary learning and extensive listening to podcasts? (2) What is the relationship between the frequency of occurrence of the target vocabulary in podcast episodes and incidental vocabulary learning? (3) What is the relationship between the distribution of occurrence of the target vocabulary across the podcast episodes and incidental vocabulary learning? The instruments applied for data acquisition were Updated Vocabulary Levels pre- and post-Tests, project-designed pre- and post- tests, weekly listening comprehension tests, digital listening journal analyses and a post-survey. The results indicate that extensive listening to podcasts leads to significant vocabulary growth.

1 Introduction

Vocabulary learning is essential in foreign language acquisition (Nation & Meara, 2010, 2020; Nunan, 2015; Schmitt, 2000). Words are the main semantic signs which convey meaning and impact the flow of communication immediately in terms of comprehension. The richness of vocabulary makes communication much easier and opens up more opportunities for EFL learners.

Traditionally, in an EFL context teachers pay due attention to vocabulary teaching which is often explicit (Nation, 2013). The vocabulary range that is taught and included in coursebooks is the most frequent vocabulary, namely the most frequent words at the 1,000, 2,000 and 3,000 levels (Nation & Meara, 2010; Schmitt, 2000). Bearing in mind that 3,000 word families (95% coverage) can be enough for basic oral

communication (Nation, 2013), an educated EFL learner, however, needs from 9,000-15,000 words for solid communication (Nation, 2006). It means that a huge range of vocabulary is left to be acquired, mostly incidentally, outside school (Nation & Chung, 2009; Schmitt, 2000). Some common ways to do so are watching movies, listening to songs, reading graded readers or novels, listening to podcasts, etc.

Incidental vocabulary acquisition has been studied in a number of contexts. It was observed when doing extensive reading of graded readers (Horst, 2005; Lee, 2007; Schmitt, 2000), novels (Pellicer-Sanchez & Schmitt, 2010) or of academic reading and writing in English medium courses (Dodigovic, 2015; Parry, 1991). Other contexts included narrow reading of texts (Abdollahi & Farvardin, 2016; Rai & Sukying, 2019). Incidental vocabulary learning has also been observed in the context of audio input such as listening to recorded lectures (Vidal, 2003, 2011) or podcasts (Mechraoui et al., 2015). There was also the comparison of a more productive manner to acquire vocabulary incidentally - reading, listening or reading while listening (Brown et al., 2008; Webb & Chang, 2015).

Despite the richness of studies on incidental vocabulary acquisition, there has not been sufficient research on how vocabulary is acquired incidentally when students are extensively exposed to podcast episodes. The former research has mainly been about how incidental vocabulary learning occurs with the written discourse accompanied with its audio representation (Webb & Chang, 2015) or with academic lectures (Vidal, 2003, 2011). To add, the only study on incidental vocabulary acquisition and listening to podcasts (Mechraoui et al., 2015) was rife with some vital questions that the researchers left unanswered.

In the current paper, the goal is to acquire data on how incidental vocabulary is impacted by listening extensively to podcasts by high school students in Armenia. The problem faced is that there is a lack of research data about how vocabulary is acquired incidentally when learners are exposed extensively to audio podcasts. Because of that gap, no solid recommendations can be made to EFL teachers on the use of podcasts for incidental vocabulary learning in Armenia, keeping in mind its increasing popularity among teenagers globally (Hasan & Hoon, 2013; Hasan & Tan, 2017; Sze, 2006).

The significance of the current mixed-method study is that it could lead to exactly such recommendations. Namely, it has the potential to answer the question whether extensive listening to certain types of podcasts can lead to a significant increase in incidental vocabulary learning. If the impact of the audio input is found to be positive among high school Armenian students with the initial knowledge of the 3,000 most frequent words, then listening to podcasts extensively for the purpose of vocabulary acquisition can be recommended to EFL teachers and students in Armenia. The formation of the habit to listen to podcasts for enriching one's vocabulary incidentally is in that case likely to assist future school graduates in increasing their EFL vocabulary independently as life-long and autonomous learners. This would be especially important for students in developing countries, who might mainly rely on self-study. While access to technology is still an issue in such contexts, podcasts, which only require a minimal technological platform and therefore are a low-cost option, would be ideally suited for low-income contexts. Moreover, it would benefit auditory learners who might be looking for an alternative to reading.

The research questions put forward in the paper are the following: (1) What is the relationship between incidental vocabulary learning and extensive listening to podcasts? (2) What is the relationship between the frequency of occurrence of the target vocabulary in podcast episodes and incidental vocabulary learning? (3) What is the relationship between the distribution of occurrence of the target vocabulary across the podcast episodes and incidental vocabulary learning?

2 Literature Review

Incidental vocabulary acquisition is one of the ways to acquire a foreign language (Ma, 2009; Pellicer-Sanchez & Schmitt, 2010; Schmitt, 2000). It is about the acquisition of vocabulary while using the language in a meaningful context (Robinson, 2001; Schmitt, 2000). An average educated foreign

language learner needs more than 9,000 words to satisfy their communicative and professional needs (Penciller-Sanchez & Schmitt, 2010). After the explicit instruction and intentional learning of the most frequent vocabulary at the 1,000, 2,000 and 3,000 word levels, incidental vocabulary acquisition seems to be the most suitable way to progress in the vocabulary of a foreign language. It requires less effort and mental focus on language by the learner, while being involved in processing information (Schmitt, 2000). Incidental vocabulary acquisition needs not only multiple exposures with a high vocabulary coverage, but also the use of noticing and guessing strategies for better retention (Nation, 2013; Schmitt, 2000; Nagy, 1997).

In the literature review, the focus will mainly be on the studies of how vocabulary has been acquired incidentally in reading. Due to its closeness to listening as a receptive skill (Mayora, 2017) most research on listening for the study of vocabulary acquisition imitates the studies with reading, particularly in terms of the conditions under which incidental vocabulary acquisition occurs. Another reason is the scarcity of research on incidental acquisition via podcasts.

Incidental vocabulary acquisition has been mostly investigated while doing extensive reading of graded readers (Cho & Krashen, 1994; Horst, 2005; Pigata & Schmitt, 2006; Schmitt, 2000; Waring & Takaki, 2003), of novels (Pellicer-Sanchez & Schmitt, 2010) and of academic texts (Dodigovic, 2015). Through reading for pleasure and content in large quantities and within their language proficiency, students in most studies have shown some vocabulary growth (Grabe & Stoller, 2002).

Graded readers have been the most popular type of input for studying how vocabulary is acquired incidentally due to the features they are designed with. Those features are the repetitive use of target vocabulary (Webb & Chang, 2015; Webb, 2007), high vocabulary coverage (Nation, 2013; Schmitt, 2000) and elaborated contexts appropriate for guessing (Elley, 1989; Schmidt, 1994). However, graded readers do not successfully incorporate mid-frequency vocabulary (3,000-9,000 levels) (Pellicer-Sánchez & Schmitt, 2010) which, sooner or later, leads to the preferences for authentic novels. However, authentic novels might be less motivating because of the number of unfamiliar words or contexts that do not contribute to guessing without dictionaries.

As for narrow reading, which is defined as "reading in only one genre, one subject matter, or the work of one author" (Cho, Ahn & Krashen, 2005, p. 58), the researchers studied incidental vocabulary learning with the topic-related input. In the case of sixty Iranian female high-school students, the materials were about global warming, money, smoking and cellphones (Abdollahi & Farvardin, 2016). Thai undergraduate students, in their turn, did the narrow reading about Buddha's life (Rai & Sukying, 2019). In both studies, the participants experienced a significant vocabulary growth which came to show that the reading of topic-related input with repeated exposure to the target vocabulary impacted incidental vocabulary acquisition positively.

Nevertheless, a point of concern with regard to extensive and narrow reading is that the distinction between narrow and extensive reading is not that clear-cut. To be specific, not every reading that "is applied to reading in only one genre, one subject matter, or the work of one author" (Cho et al., 2005, p. 58) can be unanimously defined as narrow reading.

In the current paper, students listened extensively to podcast episodes on the topic of sustainability. Extensive listening is similar to extensive reading, which happens when students read interesting materials over time with the intention to get the overall meaning without a focus on language (Richards & Long, 1987). The students in the present study listened to episodes on a weekly basis. They needed to understand the gist of the episodes without any pressure or a sense of obligation. The episodes were not connected to their curriculum. The choice of episodes from the same podcast about sustainability was to control the limitations of different genres mentioned by Webb and Chang (2015). Those differences were a reason for a "low distribution of occurrence" of the target vocabulary (Webb & Chang, 2015, p. 35). Also, because it was hard to find other podcasts with the topic and language that will be both interesting and comprehensible to high-school Armenian students, the same podcast was selected as input.

Despite the increasing popularity of podcasts in EFL contexts (Hasan & Hoon, 2013; Hasan & Tan, 2017; Sze, 2006), little research has been conducted on how vocabulary is acquired incidentally when listening to podcasts. Podcasts are a specific mode of digital information delivery, introduced to their listeners through a series of episodes. As well as audio, they can include video with animated images or real people. Most podcast channels allow subscriptions and contain Real Simple Syndication (RSS) feeds that enables a subscription to favorite podcast creators who come up with the series of episodes through which the content of a podcast is delivered.

Moreover, podcasts are delivered in oral spoken formats such as interviews, co-hosted conversations, monologues, storytelling, etc. (Burton, 2021), therefore the vocabulary range of 3,000 word families is enough to comprehend them (Nation, 2006, 2013) which turns into 6,000-7,000 word families for episodes with the focus on the academic content. As in oral communication, speech delivery in the podcast episodes is expected to be full of the most frequent vocabulary, simple sentence structures, less overloaded content and repetition (Coombe et al., 2010). Negotiation and para-linguistic features are other contextual features that are assumed to promote guessing and noticing in the current research (Jones, 2005).

In a primary study by Mechraoui et al. (2015), 17 Thai students from the experimental group at the pre-intermediate proficiency level managed to show better results in the vocabulary test after listening to podcasts than the control group of 17 students that had conventional teaching with the textbook and CD players included in the book ($p = 0.0309$). However, this study is full of unanswered questions which are essential to evaluate the research findings in terms of their reliability and validity. It is unclear whether the vocabulary in podcasts was in accordance with the learners' vocabulary size, whether the students' vocabulary size was measured or not, whether the podcasts were easy to perceive or whether they were related to the coursebook and its units in any way. Additionally, there is no information about the tests and what vocabulary was included or which was the principle for including the specific vocabulary items into the tests.

Another primary study is the one carried out by Vidal (2003) in an ESP context for 116 university students in Spain for four weeks. The researcher chose four recorded lectures that were connected to the students' discipline and shared with them. The results showed that the students both gained new vocabulary (M=30.41) and retained it (M=16.14). The instrument to measure the progress was a pre-, post- and delayed vocabulary test.

As a continuation of the previous research, in her subsequent study, Vidal (2011) had 260 university students of different proficiency levels. The aim was to compare their academic reading with their academic listening in terms of vocabulary gains. For that, she had three groups of participants: the first one read academic texts, the second listened to academic lectures and the third group received no treatment. The results showed that the students acquired the most vocabulary by reading. However, the research did not respond to the question about how images and extra-linguistic reality promoted video perception (Fortanet-Gómez & Ruiz-Madrid, 2014; Paltridge, 2006), which, in turn, could impact vocabulary acquisition (Peciller-Sanchez & Schmitt, 2010).

Gholami and Mohammadi (2015) also focused on vocabulary learning and podcasting but with podcast creation as an aim of their investigation, as well. The students who had an active integration into the podcast creation scored high in the post-test though no details in figures were provided.

As for vocabulary distribution across podcast episodes, it has been the least investigated factor in incidental vocabulary acquisition (Horst, 2005) which is defined as "the distribution of occurrences of" the target vocabulary across episodes being dependent on the number of input, which was found to be of seven episodes (Webb & Chang, 2015, p. 5). In the case of vocabulary distribution, the encounter of the target vocabulary happens across multiple episodes rather than in a single one which is the case for the study of the frequency of vocabulary. In the case of frequency in the input, what matters is the number of times the target vocabulary occurs in an episode and not the number of episodes that vocabulary is

distributed across. As suggested, the high frequency of target words makes their acquisition possible. The number of repetitions needed to acquire a word ranges from 5-16 (Cobb & Meara, 1998; Nation, 1999; Webb & Chang, 2015; Webb, 2007; Horst, 2005) to even 20 (Waring & Takaki, 2003). In parallel to the suggested figures, there is the existing argumentation that no fixed quantity of repetitions may lead to incidental vocabulary acquisition (Nation & Wang, 1999; Webb & Chang, 2015).

One of the rare studies on vocabulary distribution, as well as frequency in the input is by Webb and Chang (2015) with 10 graded readers. In their study, it was revealed that neither distribution ($r = .04$, $p = .73$), nor frequency ($r = –.03$, $p = .78$) had a strong or significant impact on incidental vocabulary acquisition. However, these findings cannot be generalized for reading graded readers extensively as they were focused on reading while listening.

Overall, none of the research questions in the current study have been put forward and answered in the context of incidental vocabulary acquisition while listening extensively to podcasts. The former studies have used a few recorded videos or have not been transparent about the research methodology and target vocabulary.

3 Methodology

3.1 Participants

The target participants were 32 Armenian high-school students (9 males and 23 females) aged 16 to 18 who demonstrated their knowledge of the 3,000-word level, selected from 78 volunteers. The proficiency levels of the students varied from pre-intermediate up to upper intermediate according to the teachers. There were 10 students in the comparison group and 22 in the treatment one. The justification of having an unequal distribution was that it was not the individuals who were being compared, but groups as a whole, in which case it is not required that the groups be equal in size.

3.2 Materials

The episodes were taken from a podcast called "Good Together" (https://brightly.eco). The podcast is devoted to raising awareness about sustainability from a number of angles, be it thrifting, eco-friendly brands, innovations, sustainability tips and tricks, etc. The channel is hosted by cofounders Liza Moiseeva (MA in Business Administration) and Laura Wittig (MA in Advanced Web and Interaction Design). Being worried about the huge amount of waste that big companies and people produce unintentionally, they started their podcast with the intention to inform and educate the average consumer about how to become sustainable. The format of their episodes is co-hosted interviews; they would invite a sustainable brand owner or someone who lives a sustainable life and ask questions about their choices, reasons for those choices and changes, in order to help their listeners become more conscious about their lifestyles. The interviews are informal, with scarce usage of terms in need of expert explanation.

For the purpose of vocabulary profiling, the episodes were initially transcribed with the program called "Otter" (https://otter.ai/), and were then checked manually.

The average length of the episodes was 40 minutes. The average number of tokens in each episode was approximately 8,000.

3.3 Instruments

With reference to the research being carried out with mixed methods, the instruments applied were both

quantitative and qualitative. The quantitative tools were the Updated Vocabulary Level pre- and post-Tests (pre- and post-UVLT) (Webb et al., 2017), project designed pre- and post-test (pre- and post-PDT), digital listening journals, listening comprehension tests (LCTs) and the post-survey. The qualitative instrument was the digital listening journal observation.

Particular attention, however, has to be paid to the pre- and post-PDT. The pre- and post-PDT was the same test with 13 multiple-choice items. Each had one-point value. The stems were not extracted from the input verbatim, but were drawn from the 1,000-3,000 word level so that the students did not face any comprehension problems.

The distracters and the key answers for the questions, in their turn, were all taken from the episodes. However, they were beyond the 3,000-word level up to the 6,000-word level with the intention to check whether the vocabulary was specifically acquired from the episodes and was unfamiliar to the students beforehand, as done in a number of similar studies on reading input (Dodigovic et al., this issue; Pellicer-Sanchez & Schmitt, 2010; Pitts, White & Krashen, 1989). Each item contained three distracters, the key answer and the "don't know" option to avoid random guessing. The transcripts of the answer keys and the distracters were provided since the input was oral.

Nevertheless, multiple-choice questions have their uniqueness in terms of giving correct answers. As mentioned by Paul et al. (1990, as cited in Nation, 2013), learners apply a number of strategies to tick the correct answer which are "knowing the answer, association, elimination, position of the options, readability of the options and guessing." (p. 542). To check the sample knowledge of the distracters, the second section of the test was developed. The students were asked to explain the meaning of each lexical unit in English or in Armenian within 20 minutes, for each correct answer being scored with 1 point, for a close or somehow close answer - 0.5 and contextually incorrect meaning - 0. All the items were with small initial letters (though still with the transcripts) including the shop names to avoid extraneous clues.

In the target vocabulary there were 54 lexical units included out of which 51 were single words, two noun phrases consisting of two decomposable constituent elements ("capsule wardrobe" and "greenhouse gas") and two fully compositional phrases/collocations to be consistent with the key in the item. The frequency and distributions of those compositional phrase/collocations were calculated separately as single words. Another decomposable phrase was a shop name that led to the choice of other shop names as distracters.

The homogeneity of the options for the single words was provided with the help of similar parts of speech belongingness and shared grammatical features. In the case of phrases, the distracters were chosen and made up in agreement with the key. It means that the phrases with two constituent elements having an Adj+Noun structure were formed. Of course, the reason behind this choice of phrases was an interest in how phrases can be acquired incidentally from the audio input which might serve as a basis for further research. It should be noted that the main concerns in forming the compositional collocations were that they would be unknown to the students and that they would have the same Adj+Noun structure so as to provide consistency or homogeneity among the options.

The vocabulary units were separated into the frequency categories used by Pellicer-Sanchez and Schmitt (2010) who used them in the context of authentic novels. The original classification underwent some modifications to fit the current research: 1-4, 5-8, 10-17 and 28 or more occurrences in episodes. Similar modifications were made for the distribution of the vocabulary across the episodes to adapt to the vocabulary included in the study by Webb & Chang (2015): 1, 2, 3, 4 and 5, 6, 7.

As for the digital listening journals, the students had to fill them in before weekly meetings by answering four questions. Their purpose was to provide information on the application of the input over the long period of incidental vocabulary acquisition (Webb & Chang, 2015) and to reveal the levels of motivation and interest participants had in the project and the topics, which are known to impact extensive listening (Nation, 2013).

Seven LCTs were assigned weekly before the topic discussions. They consisted of five multiple-choice questions. The first one was about the gist of the episode and the last four were on the details, which were mainly about practical tips. For the test, they were given six minutes. In this fashion, the episode difficulty was controlled as a factor that might have a negative effect on vocabulary acquisition (Sanchz & Schmitt, 2012; Schmitt, 2008). The tests had been piloted beforehand.

The survey was conducted in the tenth week. It consisted of 13 questions and a statement to add any ideas that had been felt to be missed in the survey. The survey skeleton was borrowed from Pellicer-Sanchez and Schmitt (2010) but was modified to the needs of the current research. The survey was conducted online within 15 minutes. The goal of the survey was to have an overall understanding of the participants' attitudes to the project, listening input and the self-assessment of their vocabulary growth. The answers were quantified into percentages.

3.4 Procedure

This quasi-experimental research was conducted in a private school in Armenia after permission was given by the school principal, community director and the parents. Participation was voluntary.

The project lasted for ten weeks because of the students' curriculum. Based on the Updated Vocabulary Levels pre-Test (pre-UVLT) participants had knowledge within the 3,000-word level. They were randomly divided into two groups, and then they took the project-designed pre-test (pre-PDT) during the first week. In the treatment group, students did their regular classes and were given one episode to listen to each week. The treatment lasted for seven weeks. There was a 30-minute meeting every week to discuss each episode and the difficulties they might have had before each meeting. Participants filled in their digital listening journals during the week and at the beginning of the meeting had a listening comprehension test. In the ninth week, they took their post-tests on different days. The project ended in the tenth week when the students completed the survey. In the comparison group, the students were exposed to no episodes and only had their regular classes which included listening to the coursebook passages and completing extra-curricular projects such as summarizing TED Talks, for instance. The coursebooks in use were Focus 3, 4 and 5 (Kay et al., 2016).

3.5 Data analysis

To answer the first research question, there was a descriptive statistical analysis of the pre- and post-UVLTs' results and the analysis of the pre- and post-PDTs' results. The results were collected from both the treatment and comparison groups. Afterwards, both paired- and independent t-test, as well as Wilcoxon analyses were conducted to understand whether the differences in the results were significant. The findings were supported by comparison with the survey responses.

The answers to the second and third research questions, which related to the frequency and distribution of the vocabulary across the episodes, entailed the usage of Spearman rho (Webb & Chang, 2015), and the gain of each word observed in the pre- and post- PDTs. One-way ANOVA was also carried out to track what impact the categories had on the learning process.

4 Results

4.1 Research question 1: Pre- and post-PDT results

As mentioned earlier, the selection of the participants was based on the pre-UVLT results, so only students with a 3,000-word level knowledge were included in the research. Then, they were separated

randomly into the treatment and comparison groups and assigned to the pre-PDT. The treatment group scored 11.05 and the comparison group 10.6 on average with the little difference in the standard deviations (SD = 2.57 for the treatment group and 3.44 for the comparison group) at pre-PDT. However, the mean was almost doubled for the treatment group in post-PDT (M=28.83) and decreased for the comparison group (M=8.9). The results of the independent t-test did not show a statistically significant difference (p = .686) in the pre-test results as shown in Table 1. However, the difference was statistically significant for the post-PDT results ($p < .001$).

Table 1

Independent Samples T-Test in the Treatment and Comparison Groups for the PDT

	T	df	P	Mean Difference	SE Difference
Pre- PDT	-0.408	30	0.686	-0.445	1.091
Post-PDT	-10.285	30	< .001	-19.986	1.943

Note: The test scales from 0 to 54

The post-PDT results indicated that due to the exposure to the extensive listening, the treatment group reached a significant vocabulary growth within the target vocabulary. The mean of the treatment group was 28.89 while that of the comparison group was 8.9 as detailed in Table 2.

Table 2

Descriptive Statistics for Vocabulary Learning Gains from Pre- and Post- PDTs for the Treatment and Comparison Groups

	Group	Mean	SD	Minimum	Maximum
Pretest	Treatment	11.05	2.57	8	15
	Comparison	10.6	3.44	6	16
Post-test	Treatment	28.89	5.97	20	39
	Comparison	8.9	1.85	7	12
Absolute gain (pre	Treatment	33.02	7.65	22.2	46.2
to post)	Comparison	-2.203	7.84	-16.65	9.26
Relative gain (pre	Treatment	41.97	11.25	26	62.5
to post)	Comparison	-4.07	11.56	-23.7	14.4

Note: The test scales from 0 to 54

Apart from the independent t-test, the paired t-test (as well as Wilcoxon signed-rank test keeping the number of the comparison group in mind) for the comparison group also revealed that the comparison group did not gain any of the lexical units in the test from anywhere else as the difference between the pre- and post-PDT results was not statistically significant (p = .221). On the contrary, Table 3 presents results showing that the treatment group experienced a significant gain in the target vocabulary ($p < .001$) after their extensive listening activities.

Table 3

Paired Samples T-Test of the Post- PDT for the Treatment and Comparison Groups

Groups	T	Df	P	Wilcoxon
Treatment	20.22	21	< .001	< .001
Comparison	1.316	9	.221	.258

Note: The test scales from 0 to 54

As for more details on the gains of the treatment group, the mean score in the treatment group from the pre-PDT was 11.05 out of 54. The score range was from eight to fifteen. After being exposed to 7 episodes within seven weeks the mean score of the treatment group became 28.89 out of 54. It denotes that within those seven weeks the students had gained 17.84 words with 25.11 words being not acquired. As a whole, most students experienced growth which ranged from 12 to 25 words.

In the current research special attention was drawn to the comprehension of the episodes. For that purpose, seven listening comprehension tests each with five questions were assigned each week during the weekly meetings. Their reliability was measured with a single-test reliability analysis in which Cronbach's alpha was 0.78, confirming the high reliability of the test. In addition, if an item was dropped it was also measured.

The episodes were proven to be easily grasped and it could be claimed that the factor of the input difficulty as an inhibiting factor was eliminated. Table 4 distinctly points out that approximately 79% of the answers were correct, on average (if 50 = 100%, then 39.5 (the mean of the overall means) = 79%). The students in the final survey also confirmed that the texts were easy to understand and contained a reasonable amount of unfamiliar vocabulary (95.5%).

Table 4

Descriptive Statistics for the LCTs Per Week

	Week1	Week2	Week3	Week4	Week5	Week6	Week7
Valid	22	22	22	22	22	22	22
Missing	0	0	0	0	0	0	0
Mean	38.6	39.09	40.46	38.18	40	38.63	41.82
Std. Deviation	7.10	6.83	7.85	7.33	6.90	7.10	7.33

Note: The test scales from 0 to 50

Moreover, the Updated Vocabulary Levels post-Test was also administered to observe the vocabulary growth at all levels. The paired-samples t-test showed that the treatment group experienced vocabulary growth at all the frequency levels: 1,000 ($p = 0.023$), 2,000 ($p = 0.010$), 3,000 ($p < .001$), 4,000 ($p = 0.014$) and 5,000 word levels ($p = 0.004$).

4.2 Research question 2: The relationship between the frequency of occurrence of the target vocabulary in the episodes and incidental vocabulary learning

One-way ANOVA was carried out to discern if the categories of the overall frequency in the input were impactful for the gains or not. The analysis showed that the role of the factor was statistically significant as $F(4, 54) = 3.167$, $p = .032$.

Afterwards, the relative gain for each word was measured because some of them had been recognized in the pre-PDT. This step was followed by the correlation analysis of the overall frequency of occurrence of each word and the relative gain of each word within the category of the overall frequency of occurrence in podcast episodes (using the formula by Webb and Chang, 2015). The correlation was proved to be statistically significant - $r = .346$, $p = .01$. Having said that, it was also detectable that the correlation was not strong, as expected, because of the recognition of some of the most frequent lexical units included in the PDT.

As indicated in Table 5, the effect of frequency in the input became most noticeable when it reached from 10 to 17 as the gains, compared to the ones within the first two categories, were double. The same

was maintained for category four which was still twice more in terms of the relative gains than it was for the first two cases. Figure 1 makes the degrees of relative gains more visible and clear.

Table 5

Frequency of Occurrence and Relative Gains

Frequency of occurrence	Number of words	Mean of relative gain from pre to post-test	Mean of absolute gain from pre to post-test
1-4 (Category 1)	21	32.7	27.8
5-8 (Category 2)	12	33.1	27.9
10-17 (Category 3)	14	65.3	33.7
28 and more (Category 4)	7	78.6	37.4

4.3 Research question 3: The relationship between the distribution of the target vocabulary across the episodes and incidental vocabulary learning

For this research question, a one-way ANOVA was conducted as well. With the help of one-way analysis of variance, the categories of distribution across podcast episodes were found out to be statistically significant for the relative gains of the target vocabulary - $F(5, 54) = 4.572$, $p = .003$.

Figure 1

Mean Relative Gains for the Frequency of Occurrence (from Pre- to Post- PDT)
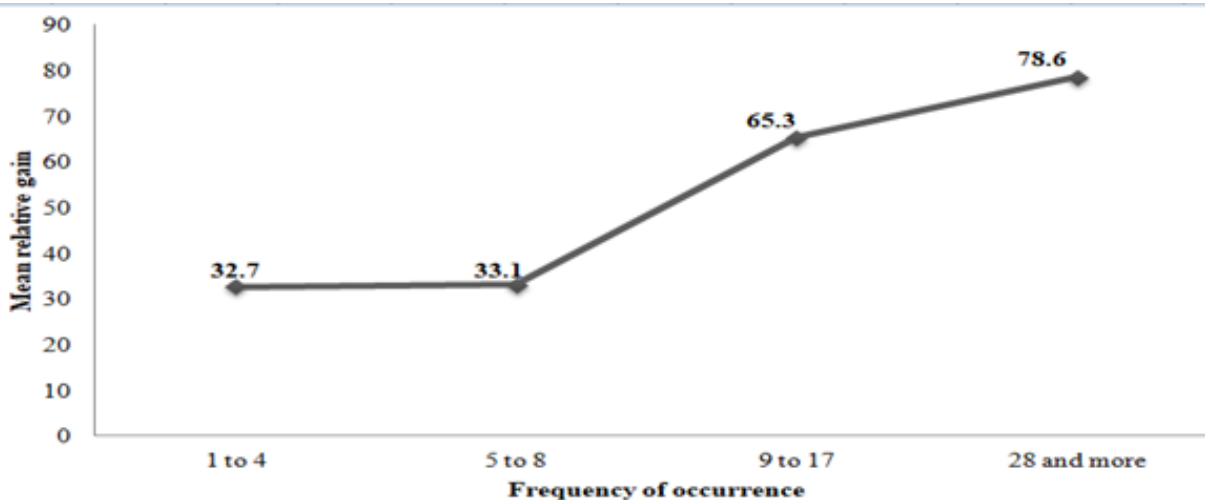
The correlation analysis with the calculation of Spearman's rho still revealed that distribution was statistically significant in the relative vocabulary gain ($r = .387$, $p = .004$). Similar to the frequency of the target vocabulary, the correlation was not strong here because of the recognition of some words in the pre-PDT.

Nevertheless, it must be mentioned that although the distribution seemed to impact the gains from distribution one to two and four up to five, six and seven quite gradually, it dropped from distribution two up to three and increased in the same amount to distribution four as shown in Table 6. In addition, the relative gain from distribution one and two increased almost twice in distribution four and five, six and seven. Figure 2 clearly indicates how substantial the gains were dependent on the distribution of occurrence across the episodes.

Table 6

Mean of Relative Gains on the Pre-to Post Test According to the Distribution of Analysis

Distribution across episodes	Number of words	Mean of relative gain from pre- to post- PDT	Mean of absolute gain from pre to post- PDT
1 (Category 1)	18	30.5	26.11
2 (Category 2)	13	35.2	28.4
3 (Category 3)	4	8.1	6.25
4 (Category 4)	14	51.26	43.57
5, 6 and 7 (Category 5)	5	60.87	28

As an extra step the correlation between the frequency of the vocabulary occurrence and its distribution across the texts was also measured which was statistically significant ($r = .566$, $p < .001$). Examples of similar lexical units were "secondhand" (19.5%), "Sheets and Giggles" (38.9%), "transparency" (27%), "ethical" (20%), "eucalyptus" (29.7%).

Figure 2

Mean Relative Gains for the Distribution of Occurrence (from Pre- to Post- PDT)
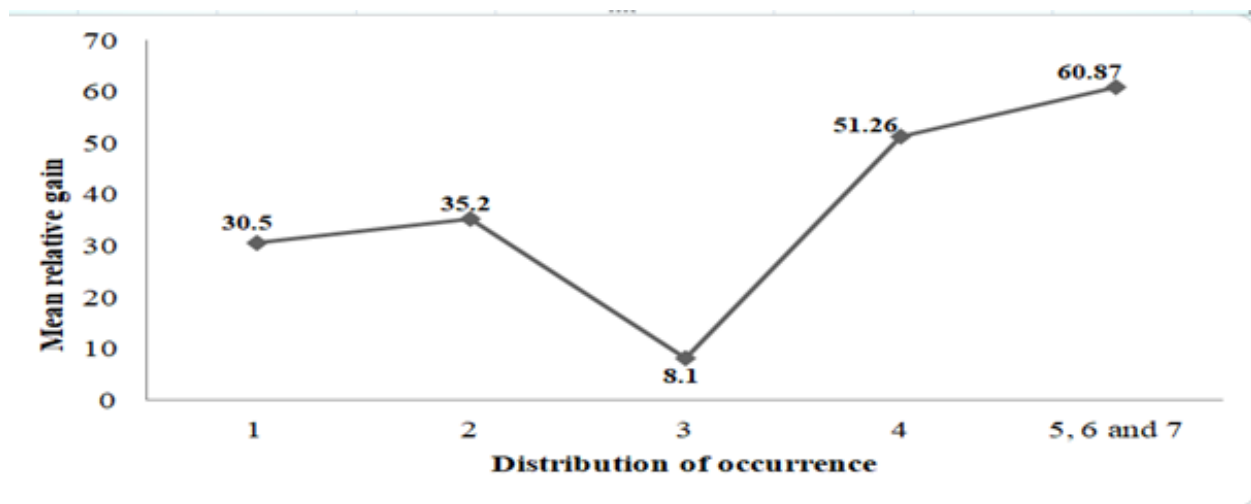
4.4 Student attitudes revealed in the survey

Overall, student attitudes were very positive. Even though 18% of the students were listening to podcasts for the first time, most students noticed the advantages of learning new vocabulary via podcasts (95%) such as learning the correct pronunciation of the word or its correct usage in the oral context. The majority confirmed that they would continue listening to podcasts to learn more words (72.7%), because of the reasonable number of unfamiliar words (86%), the use of signposts (18%), the contextually rich explanations (36%), the repetition of the vocabulary (63.6%) and humor (18%).

5 Discussion

As shown, after extensive listening to podcasts within seven weeks, students acquired 17.84 lexical units (33%) with 25.11 ones remaining not acquired out of 54 lexical units. This figure is genuinely remarkable if compared to previous results which showed gains from audio-supported reading of 19.68 words

(19.68%) within thirteen weeks (Webb & Chang, 2015). Compared to the results from pre-recorded video lectures in the study by Vidal (2003), the gains in the present research are lower (the gains were 84.5%, i.e. 30.41 out of 36 vocabulary items were acquired). The reasoning behind this difference is that in her research the duration of the exposure to the input was shorter (4 weeks), the input itself was not long either (three lectures with 14-15- minute duration) and the testing was conducted immediately after listening to the lectures.

Overall, podcasts were evaluated to be positive for incidental vocabulary learning within seven weeks (Gholami & Mohammadi, 2015; Mechraoui et al., 2015). Certainly, there could have been some forgetting of the vocabulary from the first episode till the end of the project (Webb & Chang, 2015), but as the students mentioned, the repetitive nature of the vocabulary across the episodes enhanced their ability to retain the words considerably (63.3% according to the survey) even when it seemed hard in the beginning (13%). Even more, apart from single words, there was considerable learning of some phrases such as "Sheets and Giggles" (no students knew it in the pre-PDT whereas all of the students acquired it in the end), "capsule wardrobe" (3 students had recognized it whereas the number reached 19 in the post-PDT).

One of the reasons for such gains may be the controlled difficulty of the episodes. The students found the episodes easy to comprehend, as shown by both their response to the survey question (100%) and the high scores in the LCTs (79%). In the survey, most students also indicated that they felt there was a reasonable number of unfamiliar words (86%). In this study, practicing of the vocabulary while completing the digital journals was also helpful; this was integrated into the research to control the negative effect of the research duration, following Webb and Chang (2015).

As for the impact of the frequency of the vocabulary in the input, it was shown that frequency had a significant effect on incidental vocabulary acquisition when the students listened to the episodes extensively. The findings were not in line with those by Webb and Chang (2015) in their audio-supported reading research. We believe the difference could have been due to the different learning context, individual differences, or the length of the project. The results of this study corroborated the findings by Horst et al. (1998), Pellicer-Sanchez and Schmitt (2010) and Vidal (2011) who detected a correlation in the case of graded readers and video lectures, and, importantly, revealed a significant correlation between incidental vocabulary learning and the frequency of the target vocabulary when listening extensively to podcasts. Indeed, the correlation was statistically significant ($p = .01$) but not strong ($r = .346$). The scale of the strength was connected with the recognition of some vocabulary in the pre-PDT. The most frequent lexical units that had been recognized in the pre-PDT were "awesome" (22 students), "mindset" (21 students), "hacks" (13 students). In parallel, there were two lexical units such as "Sheets and Giggles" and "ethical" that showed 100% acquisition in the post-PDT.

The difference of the relative gains between 1-4 frequency and 5-8 frequency categories was found to be small. However, the gains increased twice from the frequency of 5-4 up to 10-17. The relative gains increased for 28 and more occurrences compared to 10-17 frequency as well. As a consequence, the factor of frequency becomes of considerable effect when it reaches 10-17 and 28 and more when the input is podcast episodes. The findings were quite similar to those by Pellicer-Sanchez and Schmitt (2010) who observed the frequency of the target vocabulary in authentic novels. Similar to their research, the reason for such individual variations in the final gains might be connected to the quality of the context in which the lexical items appeared (Pellicer-Sanchez & Schmitt, 2010). Another possible explanation might be the duration of the project because of which some words were forgotten (Webb & Chang, 2015). However, further studies are necessary to explain the differences in individual gains.

In essence, the acquisition of the vocabulary happened at all the levels with the frequency of 1-4, 5-8, 10-17 and 28+ which was similar to the vocabulary acquired from extensive reading (Laufer, 2013; Nation & Wang, 1999; Waring & Takaki, 2003; Webb, 2007) but the gains significantly increased with the lexical units at the frequency of 10+ and 28+ (Pellicer-Sanchez & Schmitt, 2010; Waring and Takaki,

2003). It might be concluded that the frequency of the vocabulary in a series of podcasts does matter for greater gains.

The same held true for the distribution of the vocabulary across the episodes (r = .387, p = 0.004). Unfortunately, there has not been enough research on the concept of distribution as another form of repetition. Most of the research has been about the frequency as a manifestation of repetition which is defined as the "number of times the word is said" (Vidal, 2011, p. 61) in an episode, and so far the results have not been positive as it was the case in Webb and Chang (2015) who found no correlation between vocabulary distribution across the texts and incidental vocabulary acquisition when being exposed to graded readers (p = .73). The reasons for such negative outcomes were bound with two basic limitations such as imbalanced vocabulary distribution and the difference of the genres for the selected books (Webb & Chang, 2015). In the current paper, those limitations had been considered in the design, and the students had been exposed to the audio input with similar genres.

Consequently, the vocabulary with the highest distribution was gained more, as it happened with "transparency" whose overall frequency was 6, but the frequency in each episode was quite low (1 in episode 2, 1 in episode 3, 1 in episode 4, 1 in episode 6, 2 in episode 7) so its distribution was high, it occurred across 5 episodes out of 7. The relative gain of the word within the category of distribution was 300%. Another such example was the word "takeaway" that occurred in four episodes (episodes 1, 3, 6 and 7) just one time in each episode, but brought a 50% relative gain.

6 Conclusion

In conclusion, the present study demonstrated that extensive listening to podcasts can be effective for incidental vocabulary acquisition. The careful selection of the audio input that interested the students, that was easy in terms of comprehension and contained a balanced number of unfamiliar words allowed to arrive at significant gains in incidental vocabulary learning. The input was all of the same genre and delivered by the same content creators.

As shown, frequency and distribution of the vocabulary across episodes have a positive effect on incidental vocabulary learning. In the case of frequency in the podcast episodes, the findings for the audio-input are in agreement with previous findings for graded readers, authentic novels or listening to lectures. However, no previous studies have confirmed that the distribution of vocabulary across the input can have a positive effect on incidental vocabulary learning in graded readers, or across the input of podcast episodes. The findings in this research not only highlighted the positive impact that distribution has on incidental vocabulary learning but also indicated that the increase of distribution of the most frequent vocabulary in episodes leads to more incidental vocabulary acquisition. Furthermore, the positive attitudes of participants regarding their interest and the usefulness of the podcast activities demonstrates its great potential.

6.1 Limitations

One limitation is that the students did not have an opportunity to choose the podcasts at their own will which is supposed to happen in extensive listening/reading. The free choice of input would make it hard, if not impossible, to measure their vocabulary progress before and after the research. Another limitation is the number of lexical units which decreased after their recognition in the pre-PDT. One solution can be increasing the number of episodes. Moreover, the lack of research on the linguistic features of podcasts made it hard to present the details that were related to the research on incidental vocabulary acquisition such as podcasting as a genre, its grammar and syntactic features, typical vocabulary, etc. Additionally,

learning styles were not elaborated on in the current research which would have answered whether or not the learning style was a reason for better progress.

6.2 Recommendations for further research

One of the recommendations is to continue the research with the particular focus on the distribution of occurrence of the target vocabulary across episodes with more lexical units included. More focus on it will help to understand which categories of distribution are more beneficial to incidental vocabulary acquisition and why. That feature can be investigated not only for listening but also reading. It is suggested to include vocabulary at the 1,000, 2,000 and 3,000 levels to observe their acquisition as well.

6.3 Pedagogical implications

Relying on the current study, some implications must be highlighted. One of them is the creation of graded podcasts by keeping the results related to the vocabulary frequency and distribution across episodes in mind. With the balanced distribution of the target vocabulary throughout different episodes that is also used frequently in most episodes incidental acquisition of the target vocabulary is guaranteed. They can be applied in parallel to graded readers and explicit instruction in order to activate different learning styles to promote deep learning (Schumann, 1997).

In essence, the present study does not claim that extensive listening to podcasts is the only and the best way to acquire incidental vocabulary. Furthermore, the results are not intended to be put in comparison with the results acquired with reading. The primary goal is to suggest extensive listening to podcasts as another source of incidental vocabulary acquisition so that FL vocabulary acquisition becomes diverse, interesting and more authentic.

References

Abdollahi, M. & Farvardin, T., F. (2016). Demystifying the effect of narrow reading on EFL learners' vocabulary recall and retention. *Education Research International*, 1-10.

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*, 136–163.

Burton, L. (2021). Types of podcasts: A look at podcast formats and genres. *Masterclass*. https://www.masterclass.com/articles/types-of-podcasts-explained#what-is-a-podcast

Cho, K., Ahn, K., & Krashen, S. (2005). Reading improvement, Volume 42.

Cho, K.S., Ahn, K., & Krashen, S.D. (2005). The Effects of Narrow Reading of Authentic Texts on Interest and Reading Ability in English as a Foreign Language. *Reading Improvement*, *42*, 58.

Cho, K.S. & Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids Series: adult ESL acquisition. *Journal of Reading*, *37*, 662-7.

Coombe, C., Folse, K. & Hubley, N. (2010). *A practical guide to assessing English language learners*. The University of Michigan Press.

Dodigovic, M. (2015). How incidental is incidental vocabulary learning? In C. Gitsaki, Gobert, M., & Demirci, H. (Eds.), *Current issues in reading, writing and visual literacy: Research and practice* (pp. 203 – 215). Cambridge Scholars.

Dodigovic, M., Aghbabyan, S., S., Gevorgyan, M., Amin, M., A., Hakobyan, N., & Grigoryan, Ch. (2023). Creating a context for vocabulary growth. *International Journal of TESOL Studies*, *5*(2). https://doi.org/10.58304/ijts.20230208

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*(2), 174-187.

Fortanet-Go'mez, I. & Ruiz-Madrid, N. M[a]. (2014). A Multimodal Discourse Analysis approach to humour in conference presentations: the case of autobiographic references. *Procedia - Social and Behavioral Sciences* 173, 246 – 251.

Gholami, M., & Mohammadi, M. (2015). Podcast-mediated language learning: Levels of podcast integration and developing vocabulary knowledge. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 210-214). Research-publishing.net.

Grabe, W. & Stoller, F. (2002). *Teaching and researching reading*. Longman.

Hasan, M. & Hoon, T. (2013). Podcast applications in language learning: A review of recent studies. *English Language Teaching*, *6* (2), 128-135.

Hasan, M. M., & Tan, B. H. (2017). ESL learners' perception and attitudes towards the use of podcast in developing listening skills. *The English Teacher*, *14*(2), 160-173.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, *61*, 355–382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223.

Jones, E. (2005). Assessing students' oral proficiency. In D. Lloyd, P. Davidson, & C. Coombe (Eds.), *The fundamentals of language assessment. A practical guide for teachers in the Gulf* (pp. 75-82). TESOL Arabia Publications.

Kay, S., Jones, V. Brayshaw, D. & Trapnell, B. (2016). *Focus 4*. Pearson.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? *Canadian Modern Language Review*, *59*, 565–585.

Lee, S.Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, *38*, 150–170.

Ma, Q. (2009). *Second language vocabulary acquisition*. Peter Lang.

Mayora, C. A. (2017). Extensive listening in a Colombian university: Process, product, and perceptions. *HOW*, *24*(1), 101-121.

Mechraoui, K., Mechraoui, A. & Rafeeq, S. (2015). The effect of listening to podcast on incidental vocabulary acquisition. *International Journal of Humanities and Management Sciences*, *3*(1), 7-11.

Nagy, W.E. (1997). On the role of context in first- and second—language vocabulary learning. In Schmitt, N. and McMcarthy, M. (Eds.), *Vocabulary description, acquisition and pedagogy* (pp. 64-83). Cambridge University Press.

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

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P., & Chung, T. (2009). Teaching and testing vocabulary. In M. H. Long and C., J. Doughty (Eds.), *The handbook of language teaching* (pp. 543 - 559). Blackwell Publishing.

Nation, I.S.P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics: Second edition* (pp. 34-52). Edward Arnold.

Nation, P. & Meara, P. (2020). Vocabulary. In Schmitt, N & Rodgers, M. (Eds.), *An introduction to applied linguistics: Third edition*. Routledge.

Nation, I.S.P., & Wang, K., M., T. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*, 355–380.

Nunan, D. (2015). *Teaching English to speakers of other languages: An introduction*. Routledge.

Parry, K. (1991). Building a vocabulary through academic reading. *TESOL Quarterly*, *25*(4), 629-653.

Paltridge, B. (2006). *Discourse analysis*. Continuum.

Pellicer-Sanchez, A. & Schmitt. N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart. *Reading in a Foreign Language*, *22*(1), 31-55.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*, 1–28.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, *5*, 271–275.

Rai, P.,S. & Sukying, A. (2019). The effect of narrow reading on vocabulary learning and reading comprehension in EFL learners. *Proceedings of ICELS*, 203-217.

Richards, J. C., & Long, M. H. (1987). *Methodology in TESOL: A book of readings*. Newbury House Publishers.

Robb, T., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, *5* (2), 239-251.

Robinson, P. (2001). *Cognition and second language instruction*. Cambridge University Press.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11-26.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.

Schumann, J. (1997). The neurobiology of affect in language. Blackwell.

Sze, P. M. (2006). Developing students' listening and speaking skills through ELT podcasts. *Education Journal-Hong Kong-Chinese University of Hong Kong*, *34*(2), 115.

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*, 219–258.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*, 232–245.

Webb, S.A. & Chang, A, C-S. (2015). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667-686.

Webb, S., Sasao, Y. & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL-International Journal of Applied Linguistics*, *168*(1), 33-69.

Wodinski, M. & Nation, P. (1988). Learning from graded readers. *Reading in a foreign language*, *5*(1), 155-161.

Zahar, R., Cobb, T., & Spada, N (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, *57*, 541–572.

Nare Hakobyan, PhD, is an EFL teacher with considerable experience. She is also an international teacher trainer, curriculum designer and an EFL teaching materials developer. The fields of her research interest include Multimodality in EFL, ICT, Vocabulary Teaching and Learning, Discourse Analysis in EFL classrooms and academic settings.