ORIGINAL PAPER

Incidental Vocabulary Learning and Retention from Audiovisual Input and Factors Affecting Them

整合視聽輸入對詞彙偶發學習和記憶之影響因素

Masoud Ahrabi Fakhr[1] (ORCID) · **Davood Borzabadi Farahani**[1] ·
Ali Akbar Khomeijani Farahani[1]

Abstract

Research has recently highlighted the potential of audiovisual input for incidental vocabulary learning. However, few studies have examined the role that different factors play in incidental vocabulary learning and retention from viewing. This study aims to examine incidental vocabulary learning and retention through viewing a full-length captioned episode of an English TV program and investigate whether this learning and retention are affected by the learner-related variable of prior vocabulary knowledge and the item-related variables of frequency of occurrence, contextual clues, lexicalization, word relevance, and visual imagery. A pretest posttest-between-participant design was adopted. Data were collected from 84 Iranian EFL students who were assigned to three pairs of experimental and control groups, receiving one of the following conditions: (a) immediate, (b) 1-week delayed, or (c) 3-month delayed posttests. The experimental groups were exposed to the audiovisual input while the control groups were not. The results demonstrated that incidental vocabulary learning and retention occurred at the level of form recognition and meaning recall. Moreover, it was revealed that learning and retention were affected by frequency of occurrence, contextual clues, lexicalization, visual imagery, and prior vocabulary knowledge, with the effects of the latter three possibly persisting after 3 months.

✉ Masoud Ahrabi Fakhr
    Masoud.Ahrabi@gmail.com

[1]   Department of English Language and Literature, Faculty of Foreign Languages and Literatures, University of Tehran, Tehran 14155-6553, Iran

抽象

近幾年的研究已驗證了整合視聽訊息輸入對詞彙偶發學習的潛力。然而，很少有研究探究觀看視聽影音資料時，學習者詞彙偶發學習和記憶的影響因素及作用。本研究旨在探究學習者觀看完整一集帶有字幕之英語電視節目後，影響偶發詞彙學習和記憶之因素，包含先前的詞彙知識，詞彙的相關變數(出現頻率、語境線索、詞彙化、詞彙相關性和視覺形象)。本研究採用參與者間的前後測設計，從84名學習英語的伊朗學生收集數據，他們被分配到三對實驗組和對照組，接受以下各項測驗:(1)立即後測、(2)延遲一周後測、及(3)延遲三個月後測。三對實驗組接受了不同的視聽整合輸入，而對照組則沒有。結果表明，偶發的詞彙學習和記憶的效應最主要出現在詞彙辨識和詞彙意思之記憶。此外，結果顯示，詞彙的學習和記憶最主要受到詞彙出現頻率、語境線索、詞彙化程度、視覺形象和先前詞彙知識之影響，而後三者的影響在三個月後仍可能存在。

Introduction

Vocabulary knowledge plays a pivotal role in the comprehension of a second language (L2). Considering the limited amount of classroom time that can be spent on learning vocabulary, it is improbable that all words required for understanding spoken or written discourse can be learned within the boundaries of the classroom [47]. This can be especially problematic for English as a foreign language (EFL) learners who are not exposed to authentic L2 input in society. Many researchers suggest these learners engage in extensive reading as a potential source to consolidate their vocabulary knowledge (e.g., [6, 19, 22, 44]). Nevertheless, studies suggest EFL learners prefer to view English audiovisual media over reading outside the classroom [27, 30]. This preference toward audiovisual media, which are easily accessible through TV and the Internet in EFL settings, is a phenomenon whose potential should not be overlooked in L2 research. The significance of audiovisual input lies in its combination of moving image, sound, and even text (in the form of captions), as well as invaluable L2 socio-cultural information [39]. Hence, research on L2 acquisition has recently paid special attention to the value of audiovisual media for language learning (e.g., [3, 17, 28, 36]). Corpus-based studies have, for instance, addressed the vocabulary demands of TV programs and movies, estimating 3000 word families are required to reach 95% lexical coverage [49, 50]. Experimental studies have revealed that viewing with captions enhances comprehension [15, 18] as well as vocabulary learning [1, 15, 18] and inspires bottom-up processes leading to greater automatic word recognition [14]. Moreover, there is mounting evidence that viewing audiovisual media leads to incidental vocabulary learning [16, 31, 33, 38]. Some studies have also examined factors that affect vocabulary learning from audiovisual input [3, 29, 31, 36]. Nevertheless, there are gaps in previous research that need to be addressed. First, a shortcoming in previous studies stems from their use of short clips, film excerpts, or short educational videos [16, 33, 37, 42, 43], none of which is fully representative of what a

viewer might normally do at home [31]. Second, previous studies have only examined learning of a small number of items selected from the audiovisual materials used, something which is noted to underestimate actual learning gains [26, 31]. Third, most previous studies have not factored in retention of vocabulary (e.g., [4, 16, 29, 33]), and some have acknowledged it as a limitation [4, 16], as their results are solely restricted to recognizing or recalling words immediately after viewing the input. Further, those that have actually attempted to include retention have not produced reliable data [31, 48] due to the testing phenomenon [5]. Finally, factors that affect vocabulary learning are studied to aid materials developers to design more effective resources as well as enabling us to better perceive the extent to which different materials can contribute to vocabulary learning [3]. Nevertheless, viewing research has mostly restricted its attention to the positive effects of only two factors: frequency of occurrence and prior vocabulary knowledge (e.g., [3, 29, 36]), while evidence from reading research [2, 25, 40, 45, 46] and theoretical assumptions in viewing research [11, 24, 35, 36] indicate that there are other variables that may play substantial roles in vocabulary learning from viewing. The present study aims to address these gaps by employing a full-length captioned episode of an English TV program, examining learning and retention of a large list of target items, adopting a methodology that circumvents the risk of the testing phenomenon [5], and exploring the effects of not just two but six variables on vocabulary learning and retention.

Background

Vocabulary Learning from Audiovisual Input

The theoretical foundation of a possible superiority of audiovisual input over audio or text alone can be traced to Paivio's [24] dual-coding theory and Mayer's cognitive theory of multimedia learning [11, 12]. The central assumption in both theories is that engaging both auditory and visual channels is expected to enhance information processing, leading to an improved depth of processing and better recall [18]. In Mayer's theory, captions were initially assumed to interfere rather than facilitate learning, but L2 learning was soon excluded from this assumption when Lee and Mayer [10] conducted an experiment which resulted in acknowledging the facilitative role of captions for L2 learners. As a result, theoretically, the combination of image, audio, and text constitutes the best material for L2 learning.

In empirical research, growing evidence suggests that audiovisual input is indeed a source of incidental vocabulary uptake [3, 16, 28, 31, 33, 38, 42, 43]. Vidal [42, 43], for instance, examined gains from three 14–15-min academic lectures and targeted 12 words in each lecture. The results indicated significant incidental vocabulary uptake from pretest to posttest. Puimège and Peters [33] investigated gains in 15 words from a 30-min excerpt of a TV program and reported learning of 3.5 words in form recognition and meaning recall. While investigating different types of captioning techniques, Montero Perez et al. [16] measured knowledge of 12 common target words from three 2–5-min videos. Their study showed that the participants who watched captioned video generally performed better than the no captions group and recognized 9–12 and recalled 2–3 of the target words. Similarly, Peters [28] measured knowledge of 36 target words after viewing an approximately 11-min documentary excerpt. The results

indicated higher gains for the groups with captions compared to those with L1 subtitles and no captions, with those who viewed captioned video recognizing approximately three and recalling four words.

A common feature of these studies is their short length of input and a small number of target words. Peters and Webb [31] seems to be the only study that has employed a full-length TV program (no captions) and targeted a relatively larger list of 64 target items. They carried out two experiments with Dutch EFL learners, discovering an average gain of four words at the level of meaning recall and meaning recognition. It was additionally found that learning was influenced by prior vocabulary knowledge, frequency of occurrence, and cognateness.

Retention

Two studies that examined the effects of retention were Vidal [42, 43], both of which employed delayed posttests after 4 weeks and reported a retention rate of approximately 50% of the words learned through viewing academic lectures. Nevertheless, studies that administered their posttests with shorter delays obtained different results. Peters and Webb [31] employed 1-week delayed posttests and found out that the number of learned words actually increased in the delayed posttests, realizing that the results could not be attributed to the learning treatment. A similar problem was also detected in Webb, Newton, and Chang [50]. These results could be explained by the possibility of the testing phenomenon [5], which indicates that exposure in the immediate posttest positively affects learners' scores on the delayed posttest. Even the results of Vidal's [41, 42] studies may have been affected by the testing phenomenon, rendering the actual retention rate of learned words after 4 weeks much lower than the 50% reported.

Factors Affecting Incidental Vocabulary Learning

Incidental vocabulary learning is likely to be affected by learners' *prior vocabulary knowledge*, which is the general vocabulary size learners have before being exposed to the treatment [33], possibly because more known words in a context results in better comprehension and assigning more focus on the fewer unknown words that are left to be learned. This notion has been corroborated by previous viewing studies which have found positive correlations between prior vocabulary knowledge and incidental vocabulary gains [15, 29, 31, 33, 34], although Rodgers and Webb [36] did not find such a relationship.

Moreover, since individual words have their own learning burden [37], different item-related factors may additionally affect incidental learning. Repeated encounters with a lexical item in a text, or *frequency of occurrence*, is a factor that has received a lot of attention in the literature of incidental vocabulary learning [19]. For instance, Nation [19] argues that it is critically important for extensive reading programs to give learners the opportunity to keep meeting the words they have previously seen for incidental learning to be successful. Reading research even suggests that at least six to sixteen exposures are required for learners to develop a deep word knowledge [33]. Another factor that might play a role is *word relevance*, which is the relevance of a word to understanding a text. According to Vidal [43], words such as the technical words that are closely related to the topic of a lecture are more likely to be noticed and

learned. In viewing research, positive correlations have been found between vocabulary gains from viewing and both frequency of occurrence [29, 31, 42, 43] and word relevance [42, 43]. Nevertheless, these findings have been contrasted by Feng and Webb [3] and Peters and Webb [31], who did not find a significant role for the frequency of occurrence and word relevance, respectively. Therefore, further research to clarify the role of these factors seems warranted.

Furthermore, the presence of *visual imagery* is the factor that distinguishes audio-visual input from listening and reading and theoretically accounts for its superiority as learning material over them [11, 24, 35, 36]. Nevertheless, only one study has experimentally examined visual imagery [28]. Interestingly, Peters [28] found words with visual support to be three times more likely to be picked up by learners.

There are at least two other item-related variables that are likely to influence incidental vocabulary learning but have not been considered in previous viewing research. Studies on reading have reported significant effects for the availability of *contextual clues* (i.e., information in the context that helps infer the meaning of an unknown word) [40, 45, 46] and *lexicalization* (i.e., presence of a conventional L1 equivalent for an L2 word) [2, 25]. According to Teng [40], words that appear in rich and clear contexts provide relevant cues to word meanings and are more likely to be learned. It is also hypothesized that an unfamiliar L2 word which is lexicalized in one's L1 represents existing, or largely overlapping, semantic and syntactic information in the learner's mental lexicon, and is therefore easier to learn [2]. Research is needed to explore these assumptions in the context of audiovisual input. It is noteworthy that data on the effects of any of the mentioned factors on retention are scarce.

Research Questions

The present study attempts to fill gaps in the previous literature with regard to (a) short length of audiovisual input, (b) small number of target items, (c) lack of retention data on incidental vocabulary learning, and (d) scarcity of data regarding factors that affect incidental vocabulary learning and (especially) retention. The following questions are addressed:

1. Does watching a full-length captioned episode of an English TV program have an effect on incidental vocabulary learning and retention?
2. What is the relationship between vocabulary learning and retention through watching a full-length captioned episode of an English TV program and the following variables: prior vocabulary knowledge, frequency of occurrence, contextual clues, lexicalization, word relevance, and visual imagery?

Method

This study used a pretest-posttest between-participants design. To provide more reliable retention data, van Zeeland and Schmitt's [41] design, which used different groups for immediate and delayed posttests to circumvent the possibility of contaminating delayed posttest data, was followed. Three experimental groups were exposed to the audiovisual input, while three corresponding control groups were not. The groups were randomly assigned to one of immediate, 1-week delay (1WD), or 3-month delay (3MD)

conditions. The audiovisual material, the questionnaire, the potential target items, the tests, and the procedure were piloted with a group of language learners resembling the participants of the study.

Participants

This study was conducted with 84 Iranian participants, with Persian as their first language, in their first year of studying EFL at Farhangian University in Iran. Initially, 88 learners took part in the pretest, 87 of whom showed mastery of at least 80% of the first two levels (i.e., the first 2000 words) of the updated Vocabulary Levels Test [51] and became eligible for inclusion in the study. This precaution was taken to eliminate participants who were highly unlikely to comprehend the content of the video [50]. Three other students were eliminated from the study since two did not take part and one scored considerably lower in the posttest. A summary of the groups and participants of the study is provided in Table 1.

Audiovisual Input

The authentic TV program selected for the study was a full-length, 1-h Nova documentary titled *Why Do We Talk*. The documentary discussed issues related to language learning, hence ecologically valid for the English major participants of the study. An analysis of the documentary's lexical profile using RANGE [21] and Nation's [20] BNC/COCA word lists revealed that 92.5% of the vocabulary was from the most frequent 2000 word families, and the addition of the third level accounted for 96.6% of the documentary's vocabulary coverage. This assured us that the target material could not be too demanding for the target population [50]. Moreover, findings from the post-viewing questionnaire in the pilot test suggested most learners found the input interesting, relevant, and appropriate in terms of difficulty.

Target Words

A relatively large number of 102 words from the audiovisual input were piloted, 96 of which were selected for the study. Four words (i.e., *enormous*, *internal*, *massive*, and *unique*) were eliminated from the final test due to being known by 80% or more of the participants, and two (*colony* and *disorder*) were not included for being polysemous, as scoring them was found problematic in the meaning recall test. All target items were selected from the 2000-word level or above. All items were analyzed in terms of the

Table 1 Groups and participants

Groups	Number of participants
IE	20
IC	14
1WDE	16
1WDC	11
3MDE	11
3MDC	12

IE = immediate experimental, *IC* = immediate control, *1WDE* = 1-week delay experimental, *1WDC* = 1-week delay control, *3MDE* = 3-month delay experimental, *3MDC* = 3-month delay control

following factors: frequency of occurrence, contextual clues, lexicalization, word relevance, and visual imagery.

Frequency of occurrence was obtained through counting the number of times the target items occurred in the input. The contextual clues, lexicalization, word relevance, and visual imagery of the target items were determined through ratings of three experts with graduate degrees in English. For contextual clues, the raters considered all contexts where the target items appeared and rated the items based on a 4-point scale adopted from Webb [45]; the ratings ranged from 1 for contexts that gave participants no clues for guessing the meaning correctly to 4 for contexts that gave participants a good chance of inferring the correct meaning. For lexicalization, Chen and Truscott [2] were followed in that the items with no conventional L1 equivalent or fixed L1 item with the same meaning characteristics were considered as non-lexicalized (i.e., a rating of 0) and the items with such characteristics as lexicalized (i.e., a rating of 1). Word relevance was rated by means of a 7-point scale adopted from Peters and Webb [31]; words perceived to be more useful for understanding the input received higher ratings (i.e., 7 being very relevant to understanding the content) and vice versa (i.e., 1 being not relevant to understanding the content). Finally, we devised a scale to quantify the visual imagery of items. The raters took into consideration the degree to which an image co-occurred its aural form [35] and used the following scale to help their rating:

1. (Almost) no visual imagery that may make it possible to infer the meaning of the target word is available.
2. Visual imagery which may make it possible to infer the meaning of the target word is partly available. Participants might gain partial knowledge.
3. Visual imagery is available and gives participants a good chance of inferring the meaning correctly. Participants should gain at least partial knowledge.

The average score of the three raters was used in our analyses for contextual clues, word relevance, and visual imagery, while an agreement between all raters' judgments or at least two out of three determined whether an item was labeled as lexicalized or non-lexicalized. The target items and their variables are provided in Appendix Table 8.

Instruments

Vocabulary Knowledge Test

Learners' prior vocabulary knowledge was assessed by means of a 150-item meaning recognition test [51]. The test checks participants' vocabulary knowledge in the first to the fifth 1000-word frequency levels sourced from Nation's [20] BNC/COCA word lists and has shown reliability estimates of .96 and separation estimates of 4.72 and above.

Form Recognition and Meaning Recall Test

To test learners' knowledge of the 96 target items, a paper-and-pencil test adapted from Peters and Webb's [31] was administered. In addition to presenting the items in their written forms, the recordings of the aural forms of the items were played twice at the start of the test. The same test was used as pretest and posttest. Recordings of individual

items recorded by native speakers for dictionaries were used to enable us to scramble the order of item presentation in the posttest and therefore control for order effects. The test consisted of two parts, the first section of which focused on form recognition and the second on meaning recall. According to Peters and Webb [31], this format minimizes the test duration compared with administering two separate tests. An example of the form recognition and meaning recall test is provided in Fig. 1.

Both the pretest and posttest showed excellent reliability. The form recognition pretest had a Cronbach's alpha of .95, and the form recognition posttest had a Cronbach's alpha of .93. The meaning recall pretest had a Cronbach's alpha of .95, and the meaning recall posttest had a Cronbach's alpha of .91.

Questionnaires

A background questionnaire was given to participants asking them questions regarding their native language, their exposure to English TV compared to other sources of English input, and their use of captions while watching TV programs. Learners were also asked to complete a post-viewing questionnaire adapted from Peters and Webb [31] after viewing the program. Four 5-point scale questions helped us to ensure that participants liked the selected TV program and deemed it relevant to their field of study. One open-ended question focused on learners' comprehension of the program to verify if learners generally understood its gist. Another open-ended question asked what vocabulary items were learned from the program, enabling us to discover any gains other than the ones measured in the study. The post-viewing questionnaire was specifically helpful for the pilot study to ascertain that the right input material was chosen for the study.

Procedure

The data were collected in two sessions from the IE and all control groups and in three sessions from 1WDE and 3WDE groups. These experimental groups were told that they were going to be tested on their comprehension. The participants were debriefed about the aims of the study after all data were gathered. A summary of the procedure is presented in Table 2.

Scoring and Analysis

The data from all tests were scored dichotomously with incorrect responses receiving 0 and correct responses receiving 1. The meaning recall tests were scored by two raters, who adopted a lenient scoring procedure in both pretest and posttest; for example, they accepted *animal group* or *type* as synonyms for *species* and did not specifically look for exact definitions or equivalents. The inter-rater reliability was 98% for all tests, and the

Item	Have you ever heard the word before? Do you recognize the word?	Give the meaning (translation, English synonym, definition, …)
abruptly	Yes ☐ No ☐	
acquire	Yes ☐ No ☐	
adapt	Yes ☐ No ☐	

Fig. 1 Example of form recognition and meaning recall test

Table 2 Procedure

Groups	Two weeks before treatment: background questionnaire, vocabulary knowledge test, and pretest	Treatment session:		One week after treatment: posttest	Three months after treatment: posttest
		Treatment and post-viewing question-naire	Posttest		
IE	√	√	√		
IC	√		√		
1WDE	√	√		√	
1WDC	√			√	
3MDE	√	√			√
3MDC	√				√

IE = immediate experimental, IC = immediate control, 1WDE = 1-week delay experimental, 1WDC = 1-week delay control, 3MDE = 3-month delay experimental, 3MDC = 3-month delay control

disagreements were resolved by a discussion. A learned item was one that was not known in the pretest but known in the posttest (i.e., an absolute gain). As the data were normally distributed, an analysis of covariance (ANCOVA) with learners' prior vocabulary knowledge as a covariate was computed for all groups to answer research question one. The variances between the control and experimental groups that were explained by the treatment and the covariate were calculated [23, 32]. To take into account the varying opportunities between participants for increases in knowledge, relative gains were used in the ANCOVA [8, 31]. The following formula was used to calculate relative gains: [absolute gains/(number of target items − number of known words)] × 100. To answer research question 2, we computed generalized estimating equation (GEE) analyses to perform a repeated measures logistic regression [7] on the experimental groups. GEE analysis allows for the inclusion of participant-related variables and item-related variables in one model and is not based on total learning gains or total test scores per participant but on the number of unknown cases in the pretest that could be potentially learned in the posttest. This resulted in the combination of participant, item, and response defining for each observation a correct or incorrect score for a particular participant on a particular item. The odds ratio (exp(B)), which predicts the odds of a correct response, was calculated for each parameter. The following parameters were entered into the model as covariates: learners' prior vocabulary knowledge, frequency of occurrence, contextual clues, lexicalization, word relevance, and visual imagery. Non-significant predictors were eliminated one by one in a backward stepwise selection to achieve the final models consisting of predictors with a *p* value lower than .05. The analyses were conducted in SPSS.

Results

Questionnaires

The participants reported to have an average weekly viewing of 7.8 h, which was longer than their 5.4 h of listening and 3.2 h of reading. They also opted for watching English

audiovisual materials with captions (i.e., L2 subtitles) (3.1 h) compared with doing so without subtitles (2.9 h) or with L1 subtitles (1.8 h). In the post-viewing questionnaire, almost all items listed by learners as words they learned from the audiovisual input were among our target items, although there were a few mentions of other words as well.

Prior Vocabulary Knowledge

Results of the vocabulary knowledge test (see Table 3) indicated that all groups had mean scores showing mastery of the most frequent 1000 and 2000 words (1K > 29 and 2K > 26) and a good knowledge of the most frequent 3000 words (3K > 20.5). Three t tests were conducted to examine the homogeneity of variance between each pair of experimental and control groups in terms of prior vocabulary knowledge, revealing no significant difference between immediate groups (t = 1.06, df = 32, p = 0.30, d = 0.03), 1WD groups (t = 0.10, df = 25, p = 0.92, d = 0.001), and 3MD groups (t = − 1.60, df = 21, p = 0.12, d = 0.001). Moreover, Levene's test of homogeneity of variance indicated no difference among the three experimental groups in terms of prior vocabulary knowledge means (F (2, 44) = 0.38, p = 0.69).

Form Recognition

Table 4 provides descriptive statistics for form recognition tests in all groups. In the posttest, the participants were instructed that their recognition of items should have been based on seeing or hearing the words in contexts other than the pretest. In terms of mean absolute scores, the IE group gained 8.1 words (17.2% increase (in relative gains)), the 1WDE group gained 4.7 words (11.4% increase), and the 3MDE group gained 3.2 words (9.4% increase), while their control groups gained 2.6 (7.5% increase), 1.8 (4.2% increase), and 2.5 (6.9% increase) words, respectively. Figure 2 illustrates the absolute gains in form recognition over time; as the participants in different groups did not significantly differ, these gains may be an indicator of how learned words are retained in a week and 3-month's time. The ANCOVA showed that the relative gains were significantly higher for the experimental groups than the control groups in immediate posttest (F (1, 31) = 17.55, p = .000, η_p^2 = .36) and in 1WD posttest (F (1, 24) = 6.74, p = .016, η_p^2 = .22). The difference in relative gains

Table 3 Mean scores and standard deviations (in brackets) per group and test section

Group	Total score (max = 150)	1K (max = 30)	2K (max = 30)	3K (max = 30)	4K (max = 30)	5K (max = 30)
IE (N = 20)	110.4 (17.65)	29.95 (0.22)	26.25 (3.01)	21.15 (5.13)	19.5 (6.16)	15.05 (5.96)
IC (N = 14)	117.07 (18.80)	29.71 (047)	26.86 (3.44)	22.36 (5.02)	20.29 (5.09)	17.93 (6.60)
1WDE (N = 16)	109 (18.97)	29.94 (0.25)	26.13 (3.24)	20.56 (4.65)	18.31 (6.33)	14.06 (7.29)
1WDC (N = 11)	109.73 (16.65)	29.73 (0.65)	26 (2.76)	20.91 (4.66)	18.82 (5.12)	14.18 (4.73)
3MDE (N = 11)	123.18 (14.41)	29.91 (0.30)	27.55 (2.07)	25.09 (3.42)	22 (4.24)	17.73 (6.20)
3MDC (N = 12)	112.42 (17.54)	29.42 (090)	26.75 (3.08)	21.25 (4.73)	20.25 (4.41)	14.67 (8.06)

IE = immediate experimental, *IC* = immediate control, *1WDE* = 1-week delay experimental, *1WDC* = 1-week delay control, *3MDE* = 3-month delay experimental, *3MDC* = 3-month delay control

Table 4 Descriptive statistics for form recognition tests

| Group		Pretest	Posttest	Absolute gains	Relative gains %
IE ($N = 20$)	Mean	43.50	50.05	8.10	17.18
	SD	17.30	17.34	4.02	9.95
IC ($N = 14$)	Mean	52.71	54.29	2.64	7.85
	SD	15.06	15.91	2.24	7.53
1WDE ($N = 16$)	Mean	44.31	47.31	4.69	11.40
	SD	17.35	17.63	2.63	10.19
1WDC ($N = 11$)	Mean	48.09	49.18	1.82	4.25
	SD	14.75	15.01	1.66	3.62
3MDE ($N = 11$)	Mean	62.45	64.64	3.18	9.42
	SD	13.83	13.54	1.72	6.11
3MDC ($N = 12$)	Mean	59.58	61.33	2.50	6.87
	SD	10.63	10.29	1.78	4.35

IE = immediate experimental, *IC* = immediate control, *1WDE* = 1-week delay experimental, *1WDC* = 1-week delay control, *3MDE* = 3-month delay experimental, *3MDC* = 3-month delay control

between the 3MD groups was not found to be statistically significant (F (1, 20) = .73, p = .402, η_p^2 = .03). The analysis also indicated that prior vocabulary knowledge significantly affected learning in immediate groups (F (1, 31) = 14.81, p = .001, η_p^2 = .32) and 1WD groups (F (1, 24) = 8.65, p = .007, η_p^2 = .26), while its effect was not significant in 3MD groups (F (1, 20) = .47, p = .498, η_p^2 = .02). The treatment "watching captioned audiovisual input" accounted for 36%, 22%, and 3% of the variance in the immediate, 1WD, and 3MD posttests, while prior vocabulary knowledge explained 32%, 26%, and 2% of the variance, respectively.
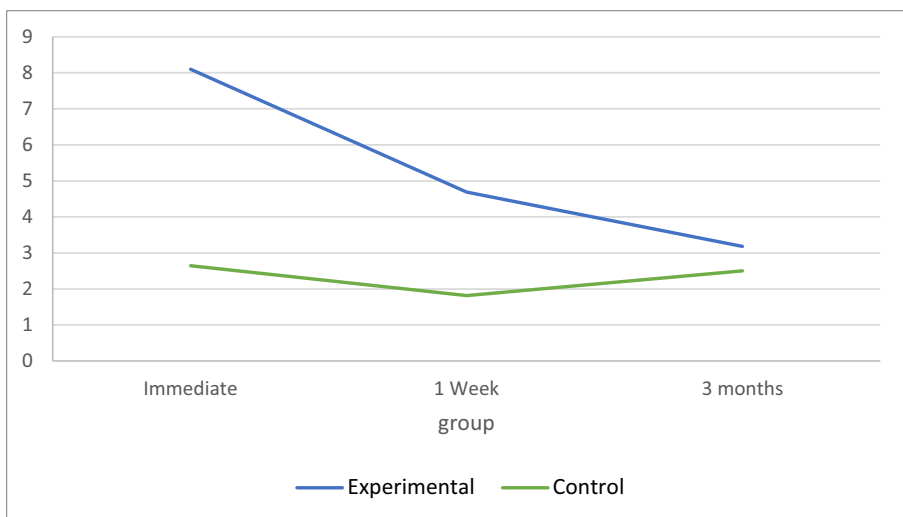
Fig. 2 Mean absolute gains for form recognition in experimental and control groups

Meaning Recall

Descriptive statistics for meaning recall tests are provided in Table 5. The participants gained 7.1 words (11.1% increase) in the IE group, 4.4 words (6.6% increase) in the 1WDE group, and 4.7 words (12.1% increase) in the 3MDE group, whereas the participants in control groups exhibited smaller gains of 1.1 (2% increase), 1.4 (2% increase), and 3.1 (5% increase) words respectively. Figure 3 demonstrates absolute gains in meaning recall over time, suggesting how retention of learned words may occur. Results provided by the ANCOVA for meaning recall were generally similar to those for form recognition. For immediate and 1WD groups, relative gains were significantly higher in the experimental groups (F (1, 31) = 36.21, p = .000, η_p^2 = .54 and F (1, 24) = 8.98, p = .006, η_p^2 = .27, respectively). No statistically significant difference was found between the 3MD groups (F (1, 20) = .84, p = .371, η_p^2 = .04). Moreover, prior vocabulary knowledge was found to significantly affect learning in all groups (i.e., F (1, 31) = 8.51, p = .007, η_p^2 = .21 in the immediate groups, F (1, 24) = 11.82, p = .002, η_p^2 = .33 in the 1WD groups, and F (1, 20) = 13.14, p = .002, η_p^2 = .4 in the 3MD groups). Therefore, "watching captioned audiovisual input" accounted for 54%, 27%, and 4% of the variance in the immediate, 1WD, and 3MD posttests, respectively, while prior vocabulary knowledge explained 21%, 33%, and 40% of this variance, respectively.

Factors Affecting Vocabulary Learning from Audiovisual Input

The GEE analyses on the meaning recall tests are reported in this paper as they provide a deeper level of vocabulary knowledge compared with form recognition. The analyses were computed on as many cases as there were items unknown to each individual in the pretest. For instance, as shown in Table 6, the IE group could potentially learn an aggregate of 1362 items; hence, the analysis of that group was computed on 1362 observations.

Table 5 Descriptive statistics for meaning recall tests

Group		Pretest	Posttest	Absolute gains	Relative gains %
IE (N = 20)	Mean	27.90	33.35	7.10	11.12
	SD	15.17	15.58	3.19	6.19
IC (N = 14)	Mean	26.71	27.64	1.07	2.02
	SD	13.60	14.83	1.54	3.24
1WDE (N = 16)	Mean	22.50	25.94	4.38	6.56
	SD	11.02	12.93	3.28	6.06
1WDC (N = 11)	Mean	26.09	27.09	1.36	1.97
	SD	11.47	11.03	1.12	1.53
3MDE (N = 11)	Mean	36.91	40.91	4.73	12.09
	SD	21.98	23.57	3.07	13.03
3MDC (N = 12)	Mean	30.25	32.67	3.08	5.00
	SD	14.32	14.67	2.75	4.43

IE = immediate experimental, *IC* = immediate control, *1WDE* = 1-week delay experimental, *1WDC* = 1-week delay control, *3MDE* = 3-month delay experimental, *3MDC* = 3-month delay control
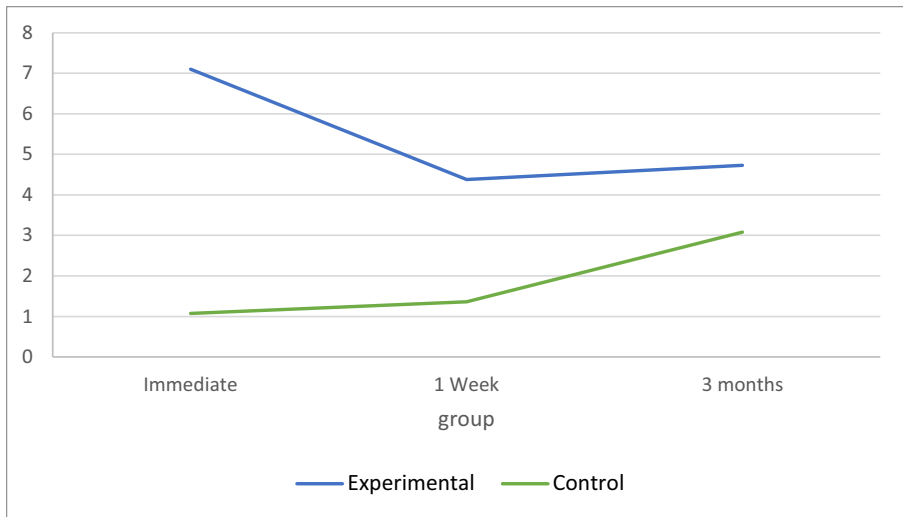
Fig. 3 Mean absolute gains for meaning recall in experimental and control groups

The results (summarized in Table 7) revealed that all factors were significantly related to immediate and 1WD learning except word relevance. In the 3MDE group, only prior vocabulary knowledge, lexicalization, and visual imagery still displayed a significant and positive correlation with the posttest scores. The odds of a correct response were 2% (in IE) to 8% (in 3MDE) higher if a participant's score on the vocabulary knowledge test increased by one. Increasing a participant's score on this test by 10 would render odds of a correct response 20% higher on the immediate, 40% higher on the 1WD, and 2.32 times higher on the 3MD posttests. When frequency of occurrence increased, the odds of a correct response were 19% higher immediately after viewing and 12% higher after a week. Five more occurrences (Exp^{5*B}) resulted in 2.35 times higher chances of a correct response in the immediate posttest and 75% higher chances in the 1WD posttest. Similar to frequency of occurrence, the effect of contextual clues decreased over time. Having the highest level of contextual clues (an increase of 3 units) increased the chances of learning 8.04 times immediately and 6.05 times after a week. Lexicalized words displayed 2.95, 3.99, and 2.34 times higher

Table 6 Number/percentage of incorrect/correct responses in meaning recall posttests

Group	Correct responses	Incorrect responses	Total responses
IE	142	1220	1362
	10.4%	89.6%	100.0%
1WDE	70	1106	1176
	6.0%	94.0%	100.0%
3MDE	53	597	650
	8.2%	91.8%	100.0%

IE = immediate experimental, *1WDE* = 1-week delay experimental, *3MDE* = 3-month delay experimental

Table 7 Gees for meaning recall posttests

Group	Variables	Wald chi-square	Sig.	B	Exp(B)
IE	(Intercept)	62.458	.000	− 7.887	.000
	Vocabulary knowledge	4.924	.026	.018	1.018
	Frequency	19.769	.000	.171	1.186
	Contextual clues	13.638	.000	.695	2.004
	Lexicalization	18.276	.000	1.082	2.951
	Visual imagery	11.739	.001	.525	1.690
1WDE	(Intercept)	36.925	.000	− 10.150	3.908E−5
	Vocabulary knowledge	10.301	.001	.034	1.034
	Frequency	3.952	.047	.112	1.119
	Contextual clues	6.917	.009	.600	1.822
	Lexicalization	4.842	.028	1.384	3.992
	Visual imagery	7.388	.007	.495	1.640
3MDE	(Intercept)	18.935	.000	− 14.134	7.270E−7
	Vocabulary knowledge	13.136	.000	.084	1.088
	Lexicalization	5.991	.014	.849	2.338
	Visual imagery	12.000	.001	.463	1.589

IE = immediate experimental, *1WDE* = 1-week delay experimental, *3MDE* = 3-month delay experimental

chances of correct responses in immediate, 1WD, and 3MD posttests, respectively. Finally, the words with clear visual imagery (an increase of 2 units) had 2.86, 2.69, and 2.52 times higher chances of being correctly answered in immediate, 1WD, and 3MD posttests, respectively.

Discussion

Incidental Vocabulary Learning and Retention from Captioned Audiovisual Input

The results of this study suggest that watching a full-length captioned TV program can result in significant vocabulary gains at the level of form recognition and meaning recall and that these learnings can be partly retained. Immediately after the viewing session, the IE group displayed absolute gains of approximately eight words in form recognition and seven words in meaning recall, corresponding to relative learning gains of 17.2% and 11.1%, respectively. This is relatively higher than learning approximately four words in meaning recall reported by Peters and Webb [31] and 3.5 words in form recognition and meaning recall found by Puimège and Peters [33]. The larger absolute gains in our study can be associated with its length of input, which was the same as Peters and Webb [31] but twice longer than Puimège and Peters [33], and the larger number of target items, as both of those studies acknowledged their limited target items as a possible cause for an underestimation of learning. Moreover, the availability of captions in our input could be a strong predictor of higher vocabulary gains, as there is

mounting evidence supporting the facilitative role of captions in vocabulary acquisition [1, 15, 18]. Overall, our results are in accordance with findings from previous research indicating that there is a potential for incidental vocabulary learning from one viewing session [3, 31, 33, 38].

In delayed posttests, the 1WDE group revealed absolute gains of 4.7 words (11.4% increase) in form recognition and 4.4 words (6.6% increase) in meaning recall. The 3MDE group showed relatively similar results of 3.2 words (9.4% increase) in form recognition and 4.7 (12.1% increase) in meaning recall, yet compared with their control groups, only the 1WDE group exhibited statistically significant gains. These findings suggest that retention of some initial learnings are readily discernible after a week but not so after 3 months. We concur with Mickan, McQueen, and Lemhöfer [13] that in experimental terms, the persistence of effects for a whole week is quite remarkable and that insignificance of the experimentally induced interference over time is "not necessarily because it is not long-lasting, but instead because of additional interference" (p. 4). If we follow studies such as van Zeeland and Schmitt [41] and Vidal [42, 43] by considering gains in delayed posttests solely as retention of learned words, our findings support theirs in that as large as half of the initial learnings seem to be retained afterwards. However, the gains observed in the control groups indicate that much of the learning observed in delayed posttests, especially in the 3MDE group, is explained by reasons other than the treatment, such as exposure to the target items in the pretest and in other sources of learning within the period before the delayed posttests.

Factors Affecting Vocabulary Learning from Audiovisual Input

The item-related variables of frequency of occurrence, contextual clues, lexicalization, and visual imagery, and the learner-related variable of prior vocabulary knowledge were all found to play a role in incidental vocabulary learning and 1-week retention, with the effects of the latter three persisting after 3 months. No significant relationship was found between word relevance and incidental vocabulary learning.

First, our results provide further support for previous studies which revealed a positive relationship between vocabulary learning and learners' vocabulary knowledge [15, 29, 31, 33, 34] and contrast Rodgers and Webb [36], who found no significant correlation between the two factors. Moreover, our results suggest that this relationship holds over time, interestingly getting stronger over longer periods. This increase can be explained by the general privilege of learners with larger vocabulary sizes to learn the target words, not solely from the treatment but from any other sources to which they were exposed before the delayed posttests. The findings substantiate that "the more words learners know, the more likely they are to pick up new words incidentally" [33, p.11].

Second, more frequent encounters with a word were found to raise the chances of learning, which corroborates similar findings in previous viewing studies [29, 31, 42, 43] and contrasts Feng and Webb's [3] results. We found the odds of learning a word were 19% higher with every increase in frequency, which closely resembles Peters and Webb's [31] report of 20% higher chances. Nevertheless, this effect weakened after a week and was insignificant after 3 months, while lexicalization, visual imagery, and prior vocabulary knowledge showed more long-lasting effects. This finding lends support to Vidal [42, 43] and Peters and Webb's [31] findings that showed frequency might not be the most important predictor in audiovisual input.

Third, a factor that was not studied in previous viewing research, but was found to be as important as frequency, if not more important, was contextual clues. Similar to frequency, contextual clues had a positive relationship with learning in the immediate and 1WD posttests. According to Teng [40], the repetition of words in more informative contexts has a larger effect on vocabulary learning than the repetition of words in less informative contexts. In fact, the absence of contextual (and visual) clues might lead learners to ignore unknown words [9, 40], regardless of their frequency. Our results are in line with reading studies espousing the facilitating role of contextual clues in vocabulary acquisition [40, 45, 46].

Fourth, lexicalization was another unexamined factor in viewing studies which was revealed to play a role in vocabulary learning immediately and over time. Lexicalized words were approximately three times as likely to be learned, supporting the hypothesis that non-lexicalized words present special difficulties for incidental learning [2, 25]. Nevertheless, our results are contrary to Chen and Truscott's [2] finding of no lasting effects for lexicalization. As lexicalization is an inherent feature in a word for a given learner, as opposed to factors which are determined by the treatment used, it seems sensible that the learning difficulty posed by non-lexicalized words would still be observable in delayed posttests.

Fifth, visual imagery was positively correlated with incidental vocabulary learning and retention. Our results approximate Peters [28], who found words with imagery to be almost three times more likely to be learned immediately after viewing and adds to the existing literature that retention of such words is more than 2.5 times more likely. The strong correlation found in this study and Peters [28] for imagery may be partly due to using documentary as audiovisual input, as Rodgers [35] observed that imagery occurs in a more contiguous fashion in documentary programs. Nevertheless, our analyses offer evidence for the assumption in previous viewing research that on-screen images in audiovisual input help learners acquire vocabulary through providing semantic support [31, 35, 36]. It is worth pointing out that while the long-lasting effects found for prior vocabulary knowledge and lexicalization may be partly explained by reasons other than the treatment, visual imagery of words is solely identified by features of the audiovisual input used in this study, making its facilitative effect on retention relatively more reliable.

Finally, our analyses indicated no significant role for word relevance. This is in line with Peters and Webb [31], whose operationalization of this factor through a 7-point scale was followed in our study, and is contrary to Vidal [42, 43], who reported larger learning gains for the technical vocabulary in lectures. As Peters and Webb [31] suggested, the difference in findings might be due to input modality, as in Vidal's studies on lectures, visual imagery might not have played a considerable role. Moreover, an academic lecture may focus on some technical vocabulary related to its topic and provide elaboration (i.e., contextual clues, a variable unexamined by Vidal) on them. The documentary program used in our study was not centered on certain technical words, as is observable in the ratings given to our target words. Overall, it seems that the type of operationalization, nature of input, and the variables examined might affect the relative significance of word relevance in different studies.

Limitations and Suggestions for Further Research

It should be noted that the generalizability of the findings of this study is limited as we had a small sample size and our participants were English majors, who may not represent the

general population of language learners. They were also used to watching (captioned) audiovisual media in their spare time, as our questionnaire results suggest. Moreover, although we used different groups for delayed posttests to circumvent the testing phenomenon, we still could not control participants' exposure to the target items in the pretest and in other sources of learning during the delay time. Using pseudowords as target items might help establish more control in future studies. Another point to consider is that in the meaning recall test, a lenient dichotomous scoring was adopted, which cannot fully account for different levels of learning. Furthermore, the type of input chosen in this study may also impose some limitations. Firstly, as we used a captioned program in this study, the degree of learning and the effects of the variables examined cannot be simply generalized to subtitled or captionless audiovisual input. Future research can examine whether the effects of different variables vary according to types of captioning. Second, the documentary genre used in this study was more likely to provide supportive imagery compared to narratives [35]. Further studies can compare the effect of imagery in different television genres. Finally, it is worth pointing out that personal variables other than vocabulary knowledge may also play a role in incidental vocabulary learning from audiovisual input. The gains in our immediate group, for instance, ranged from 0 to 13 words in absolute terms, suggesting that learning differences among the participants may be rooted in personal factors unaccounted for in our study. Hence, it seems sensible for future studies to factor in other learner variables, such as personal learning styles, learning strategies, and language aptitude in their designs.

Conclusions and Implications

The purpose of this article was to examine incidental vocabulary learning and retention through watching captioned audiovisual input. Our results provide evidence of vocabulary learning and retention from exposure to captioned audiovisual input at the level of form recognition and meaning recall. We found that learning gains were affected by participants' vocabulary knowledge and items' frequency of occurrence, contextual clues, lexicalization, and visual imagery. The results of our questionnaire, in accord with those of previous research [27, 33], suggest that EFL learners extensively watch audiovisual media and that they do this more than reading. This study suggests that there is a potential in this inclination and watching can indeed be "an effective method of learning vocabulary" [50, p. 356]. Thus, while many programs have been traditionally designed to promote vocabulary learning through extensive reading [3], our findings imply that the potential and popularity of audiovisual media for vocabulary learning should be further exploited in EFL programs. Moreover, our findings regarding factors that affect vocabulary learning from viewing can help in the selection of rich authentic audiovisual input as well as the development of educational videos for language learners.

Data Availability All data and analyses are publically available at https://osf.io/yk4v5.

Compliance with Ethical Standards

Conflict of Interest   The authors declare that they have no conflict of interest.

Appendix 1

Table 8 Target items with their frequency of occurrence, contextual clues, lexicalization, relevance, and visual imagery

Item	Frequency	Contextual clues (1–4)	Lexicalization (0–1)	Relevance (1–7)	Visual imagery (1–3)
1. Barely	1	2.67	0	3	1
2. Precise	6	2.67	1	2.33	1
3. Sequence	3	1.7	1	2.33	1
4. Trick	1	1.33	1	1.33	1
5. Distinguish	1	2.33	1	4	1
6. Species	3	3	1	4.33	1
7. Sophisticated	1	2.33	1	3	1
8. Minimal	1	2.67	1	2.67	1
9. Miraculous	1	2	1	1.33	1
10. Despite	8	3.67	1	4.33	1
11. Evolve	2	2.33	1	4.7	1
12. Fraction	1	1.7	1	1.33	1
13. Extraordinary	1	2	1	2.33	1
14. Guinea pig	2	1.67	1	2.67	1
15. Longitudinal	1	2	0	3	1
16. Ambitious	1	1.33	0	3	1
17. Unintelligible	1	2.67	1	3.33	1
18. Babble	2	3.33	1	4.67	1
19. Raw	3	3	1	4	1
20. Annotation	2	1.67	0	2.67	1
21. Insight	1	2	1	2.67	1
22. Footage	2	1.67	0	3.33	1
23. Intricate	1	1	1	1.67	1
24. Instinct	1	2	1	3.33	1
25. Blossom	2	2.67	1	3	1
26. Bury	1	2.33	1	2.33	1
27. Remarkable	4	2.67	1	2.67	1
28. Feat	1	1.67	1	2.33	1
29. Exclusively	1	2.67	1	3.67	1
30. Resemble	2	2.33	1	3.33	1
31. Chimp	3	2.33	1	5.33	3
32. Imitate	8	2.67	1	5	1
33. Vocal tract	4	3	1	5.33	3
34. Larynx	7	2.67	1	5.67	3
35. Astounding	1	1.67	1	2	1
36. Bark	3	3	1	4.67	2.33

Table 8 (continued)

Item	Frequency	Contextual clues (1–4)	Lexicalization (0–1)	Relevance (1–7)	Visual imagery (1–3)
37. Descent	1	1	1	4.33	2.67
38. Constraint	1	2.33	1	4	1
39. Vocalize	4	2.67	1	5	2
40. Elimination	1	1	1	2.33	1
41. Utter	1	2.67	1	4.33	1
42. Diagnose	2	2.33	1	4.33	1
43. Defect	3	3	1	4	1
44. Paw	1	3	1	4.33	3
45. Bowl	1	2.33	1	3.33	2.67
46. Breakthrough	2	2	0	3.33	1
47. Identify	4	2.67	1	4.33	1
48. Pacifier	1	1.67	1	3.67	3
49. Generate	1	2.33	1	3.33	1
50. Innate	11	3.33	1	6.33	1
51. Feral	1	4	0	4.67	1
52. Croak(n)	1	3.67	0	4.33	1
53. Mate	2	3.67	1	5	1
54. Yuk	1	2.67	1	4.67	1.67
55. Desperate	1	2.33	1	3.67	1
56. Isolate	8	3.67	1	5.67	1
57. Prolong	1	2.67	1	3.33	1
58. Finch	5	2.33	1	5.33	3
59. Encode	1	1.67	1	2.67	1
60. Needle	1	1	1	1	1
61. Haystack	1	1	0	1	1
62. Component	4	3	1	4.67	1
63. Multitude	1	1.67	1	2.33	1
64. Impression	2	2.33	1	3	1
65. Inherit	2	3.33	1	5.33	1
66. Faulty	1	2.33	1	3.67	1
67. Cognitive	2	1.67	1	4	1
68. Vertebrate	1	1.67	1	2.67	1
69. Coincide	2	2.33	1	3.67	1
70. Manipulate	1	2.33	0	3.67	1
71. Prime (v)	1	2	0	3.33	1
72. Enrich	1	2.33	1	2	1
73. Speculation	1	2.67	1	3.33	1
74. Proliferation	1	2.33	1	4	1
75. Explosion	1	2	1	3.33	1
76. Archeological	1	1.67	1	3.33	1

Table 8 (continued)

Item	Frequency	Contextual clues (1–4)	Lexicalization (0–1)	Relevance (1–7)	Visual imagery (1–3)
77. Hut	2	2.67	1	3.67	1
78. Abruptly	1	2.33	1	3.33	1
79. Instigate	1	2	0	3	1
80. Boom	1	2.33	1	3.33	1
81. Glue(n)	1	1.67	1	3.67	1
82. Gather	2	3	1	4.33	1
83. Spear	1	1	1	3.33	1
84. Regulate	1	2	1	3	1
85. Adapt	3	3.33	1	4.76	1
86. Unconscious	3	3.33	1	5.33	1
87. Transmission	1	1	1	3.33	1
88. Alien	2	2.33	1	4.67	1
89. Devise	1	2.67	1	3.67	1
90. Entity	1	1.67	1	2.67	1
91. Glimpse	1	1.33	0	2.67	1
92. Upbringing	1	2.33	1	4	1
93. Consensus	1	2	1	3.33	1
94. Monolithic	1	1.67	0	3.33	1
95. Acquire	2	3.67	1	4	1
96. Fully fledged	2	2.33	0	3.33	1

References

1. Chen, Y. R., Liu, Y. T., & Todd, A. G. (2018). Transient but effective? Captioning and adolescent EFL learners' spoken vocabulary acquisition. *English Teaching & Learning, 42*(1), 25–56.
2. Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. https://doi.org/10.1093/applin/amq031.
3. Feng, Y., & Webb, S. (2019). Learning vocabulary through reading, listening, and viewing: which mode of input is most effective? *Studies in Second Language Acquisition, 42*, 1–25.
4. Frumuselu, A. D., De Maeyer, S., Donche, V., & Gutierrez Colon-Plana, M. M. (2015). Television series inside the EFL classroom: bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education, 32*, 107–117. https://doi.org/10.1016/j.linged.2015.10.001.
5. Glover, J. A. (1989). The "testing" phenomenon: not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392.
6. Grabe, W., & Stoller, F. (2002). *Teaching and researching reading.* Harlow: Longman.
7. Hardin, J. W., & Hilbe, J. M. (2003). *Residual, diagnostic, and testing. Generalized estimating equations* (pp. 137–179). Boca Raton: Chapman & Hall.
8. Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207–223.
9. Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*, 327–339.

10. Lee, H., & Mayer, R. E. (2018). Fostering learning from instructional video in a second language. *Applied Cognitive Psychology, 32*(5), 648–654.
11. Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.
12. Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.
13. Mickan, A., McQueen, J. M., & Lemhöfer, K. (2019). Bridging the gap between second language acquisition research and memory science: the case of foreign language attrition. *Frontiers in Human Neuroscience, 13*, 397.
14. Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One, 4*(11), 1–5.
15. Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141 Retrieved from http://llt.msu.edu/issues/february2014/monteroperezetal.pdf.
16. Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: the effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1–2), 1–26.
17. Montero Perez, M., & Rodgers, M. P. H. (Eds.) (2019). Video and language learning [special issue]. *The Language Learning Journal, 47*, 403–518.
18. Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: a meta-analysis. *System, 41*, 720–739. https://doi.org/10.1016/j.system.2013.07.013.
19. Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
20. Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation
21. Nation, I.S.P., & Heatley, A. (2002). Range: a program for the analysis of vocabulary in texts [software]. Downloadable from http://www.victoria.ac.nz/lals/staff/paulnation/nation.aspx.
22. Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.
23. Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: a cautionary review and guide to more appropriate usage. *Second Language Research, 34*(2), 257-271.
24. Paivio, A. (1990). *Mental representations: a dual coding approach*. New York: Oxford University Press.
25. Paribakht, T. S. (2005). The influence of first language lexicalization on second language lexical inferencing: a study of Farsi-speaking learners of English as a foreign language. *Language Learning, 55*(4), 701–748.
26. Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: do things fall apart? *Reading in a Foreign Language, 22*(1), 31–55.
27. Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL-International Journal of Applied Linguistics, 169*(1).
28. Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly, 53*(4), 1008–1032.
29. Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: the differential effect of L1 subtitles and captions. *System, 63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002.
30. Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning, 69*(3), 747–782.
31. Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*, 551–577. https://doi.org/10.1017/S0272263117000407.
32. Plonsky, L. (2015). Statistical power, p values, descriptive statistics, and effect sizes: a "back-to-basics" approach to advancing quantitative methods in L2 research. In *Advancing quantitative methods in second language research* (pp. 23-45). Routledge.
33. Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*, 424–438. https://doi.org/10.1080/09571736.2019.1638630.
34. Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *The Language Learning Journal, 47*, 479–496. https://doi.org/10.1080/09571736.2019.1616806.
35. Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: the relationship between on-screen imagery and vocabulary. *ITL-International Journal of Applied Linguistics, 169*(1) Advance online publication. https://doi.org/10.5054/tq.2011.268062.

36. Rodgers, M. P., & Webb, S. (2019). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics*. https://doi.org/10.1075/itl.18034.rod

37. Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.

38. Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*(2), 50–73.

39. Talaván, N. (2007). Learning vocabulary through authentic video and subtitles. *Tesol-Spain Newsletter, 31*, 5–8.

40. Teng, F. (2016). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal, 47*(2), 145–158.

41. van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: a dimensions approach. *System, 41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012.

42. Vidal, K. (2003). Academic listening: a source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56–89. https://doi.org/10.1093/applin/24.1.56.

43. Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x.

44. Wang, Y. (2013). Incidental vocabulary learning through extensive reading: a case of lower-level EFL Taiwanese learners. *Journal of Asia TEFL, 10*(3).

45. Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65.

46. Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232–245.

47. Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

48. Webb, S., Newton, J., & Chang, A. C. S. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x.

49. Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics, 30*, 407–427. https://doi.org/10.1093/applin/amp010.

50. Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x.

51. Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test: developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69.