System 103 (2021) 102650

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Incidental lexical mining in task repetition: The role of input, input repetition and individual differences

P.T. Duong [a,b,*], M. Montero Perez [c], L.Q. Nguyen [d], P. Desmet [a,b], E. Peters [a]

[a] KU Leuven, Department of Linguistics, Belgium
[b] KU Leuven, imec research group itec, Belgium
[c] Ghent University, Belgium
[d] FPT University, Department of English, Ho Chi Minh City, Vietnam

A R T I C L E I N F O

Keywords:
Lexical mining
Incidental vocabulary learning
Input repetition
Working memory
Prior vocabulary knowledge

1. Introduction

Input plays an essential role in L2 learners' lexical development. Research has found that learners can incidentally learn new vocabulary while reading (e.g., Pellicer-Sánchez & Schmitt, 2010), listening (e.g., Jin & Webb, 2020), reading-while-listening (e.g., Webb, Newton, & Chang, 2013), and TV viewing (e.g., Peters & Webb, 2018). Despite the growing number of studies on vocabulary acquisition, few have explored L2 learners' vocabulary use in output after being exposed to a particular input type. Samuda (2001) coined the term *mining of input* to refer to learners' ability to use language elements borrowed from input, that is, "mined from input" in an output activity without being instructed to do so. A few recent studies have focused on learners' ability to mine lexical items from input in speaking/writing (e.g., Hoang & Boers, 2016; Nguyen & Boers, 2018; Yang, Shintani, Li, & Zhang, 2017). However, these studies have mainly focused on L2 learners' mining of single words in an immediate output task and not in a repeat task.

Research has documented a positive effect of task repetition on some aspects of lexical use (e.g., lexical sophistication, lexical

* Corresponding author. KU Leuven, Department of Linguistics, Belgium.
E-mail addresses: phuongthao.duong@kuleuven.be, duongthao204@gmail.com (P.T. Duong), maribel.monteroperez@ugent.be (M. Montero Perez), quocnl2@fe.edu.vn (L.Q. Nguyen), piet.desmet@kuleuven.be (P. Desmet), elke.peters@kuleuven.be (E. Peters).

https://doi.org/10.1016/j.system.2021.102650
Received 22 February 2021; Received in revised form 14 September 2021; Accepted 15 September 2021
Available online 17 September 2021

diversity) in repeat output tasks (Kim, Crossley, Yung, Kyle, & Kang, 2018). However, we are barely aware of the effect of task repetition on lexical mining. Lynch (2018) claimed that learners perform better in a repeat task if they engage in cognitive activities related to the first output task after the first task performance. Nguyen and Boers (2018) suggested exposing L2 learners to the same input before and after their output performance as a kind of cognitive activity to provoke their reflection on the output. However, no study has investigated to what extent such form of input repetition can influence lexical mining.

To fill these gaps, the present study aimed to (a) examine L2 learners' lexical mining in an immediate and repeat output task and (b) further explore to what extent repeating the input after the immediate output task influences learners' lexical mining in the repeat output task. This study also seeks to understand the roles of prior vocabulary knowledge and working memory as mediating learner-related variables since these variables have been found predictive of incidental vocabulary learning (Montero Perez, 2020).

2. Background

2.1. Incidental L2 lexical mining

Research on L2 lexical mining is scarce. Boston (2008) set the stage by comparing L2 learners' lexical mining from written input (i.e., written instructions) and spoken input (i.e., pre-task recordings) in follow-up conversations. He found that learners could mine words from the written instructions, but they could not from the recordings. In contrast, Hoang and Boers (2016) demonstrated that lexical mining from spoken input can occur. They found that L2 Vietnamese learners could mine a substantial proportion of single words (68%), but only a limited number of formulaic sequences (approximately 6%) from the spoken input (i.e., an L2-captioned audio-recorded story) in an immediate storytelling task. While Hoang and Boers counted the total number words shared between the input and the output, other studies only focused on the mining of target words (i.e., words that are unknown to learners as evidenced in a pretest) (e.g., Nguyen & Boers, 2018; Yang et al., 2017). Yang et al. (2017) found that learners did not mine any target words from the reading texts in their essays. However, Nguyen and Boers (2018) showed that learners could mine several ones (i.e., 5 out of 18 words) from the spoken input (i.e., a TED Talk with no L2 caption) in an immediate oral summary task. In short, findings regarding learners' lexical mining seem to vary across studies, possibly because of the differences in task design and lexical mining measures. Further, these studies have mainly focused on single words in immediate output tasks. Few studies have explored the mining of formulaic sequences except for Hoang and Boers (2016).

2.2. The role of task repetition and input repetition

It is not uncommon for learners to perform an output task more than once. In the case studies presented by Lynch and Maclean (2000), students performed repeat output tasks by presenting their posters to different visitors. While the task content is repeated, language performance (e.g., lexical use) per time of repetition tends to vary (Bygate, 2018). A notable example is a study by Kim et al. (2018) who found that L2 learners produced more sophisticated words (i.e., words that are less frequent, less familiar, and with higher age of acquisition) as they repeated the same oral task. Fukuta (2016) explored lexical diversity (i.e., how many different words are used in a text) and found that learners made efforts to widen their word choice in the repeat performance. In short, it seems that task repetition may help learners to retrieve more sophisticated and diverse words from their lexicon. However, it remains unclear how task repetition affects learners' lexical mining, that is their ability to use words from input rather than from their interlanguage system.

Lynch's (2018) review pointed out that L2 learners might perform better in the repeat task if they engage in cognitive activities related to the first output task (e.g., watching (non) native speaker's performance of the same task, correcting the first output task's transcription) right after the first performance. Lynch argued that these activities might help learners reflect on their output and prompt them to perform better in the repeat performance. Of relevance might be a recent study by Nguyen and Boers (2018) who suggested that exposing learners to the same input text (i.e., a TED Talk video) after they performed an output task (i.e., orally summarize a TED Talk video) (i.e., input-output-input) could be useful for output reflection. Specifically, Nguyen and Boers argued that repeated input might trigger learners' need to make a comparison between their use of newly encountered words in the output and the use of those words in the repeated input. In this respect, we argue that even though the input-output-cycle was not intended to foster vocabulary use during repeat performance but rather vocabulary acquisition, its cognitive mechanisms may have prompted learners to reflect on their output. As the reflection element is deemed useful for learners to perform better in the repeat oral performance (Lynch, 2018), it is possible that this cycle might benefit lexical mining in the repeat performance as well. Yet, further research is warranted to justify this hypothesis.

2.3. Prior vocabulary knowledge and working memory

The role of prior vocabulary knowledge has been widely acknowledged in incidental vocabulary learning research (for a review, see Peters, 2020). Prior vocabulary knowledge involves receptive (i.e., knowing the word meaning) and productive knowledge (i.e., producing the form from a given meaning), which can be estimated by different types of tests (for a review, see Read, 2020). It has been shown that learners' vocabulary knowledge is positively linked to their output production. However, most studies have used receptive vocabulary tests (Milton, 2013). Only a few studies have employed productive vocabulary tests to investigate the relationship between learners' vocabulary knowledge and their spoken output (e.g., De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Noreillie, Desmet, & Peters, 2020; Uchihara & Saito, 2019). While productive vocabulary knowledge has been found to be related to various aspects of learners' spoken output (e.g., fluency, general speaking proficiency), the association between receptive vocabulary knowledge and L2 learners' output remains unclear (for a review, see Uchihara & Clenton, 2018). To the best of our knowledge, no studies have examined the role of learners' receptive and productive vocabulary knowledge in lexical mining in spoken output tasks.

In addition to prior vocabulary knowledge, working memory has also received attention from SLA researchers. Baddeley's (2003) model states that working memory (WM) involves a *domain-general central executive system*, which relates to attention, and two domain-specific storage systems: *phonological working memory* for storing and handling auditory information, and *visuo-spatial working memory* for maintaining and manipulating visual and spatial information. A positive relationship has been found between WM and L2 vocabulary gains but mainly at the receptive knowledge aspects (e.g., form recognition) (Montero Perez, 2020). If we treat lexical mining as a form of productive retrieval, it remains unclear to what extent learners' WM will influence lexical mining.

In brief, the literature review has pointed to a number of gaps. Findings regarding lexical mining in immediate output tasks remain ambiguous. No evidence has been established on the difference in lexical mining between an immediate and a repeat output task. Further, no studies have explored the influence of repeating input after an immediate output task (i.e., input repetition) on lexical mining in a repeat output task. Also, learner-related variables (e.g., prior vocabulary knowledge and working memory) have not been examined in previous lexical mining studies. To address these issues, the following research questions were formulated:

1. What is the difference between lexical mining in an immediate and in a repeat output task?
2. What is the effect of input repetition on L2 lexical mining in the repeat output task?
3. To what extent are L2 learners' prior vocabulary knowledge and working memory related to their lexical mining in the immediate and the repeat output tasks?

3. Methodology

3.1. Participants

The participants were ninety-one L2 learners of English[1] (aged 19–21) (50 females, 41 males) who were native Vietnamese undergraduates. Their English proficiency was expected to be at A2-B1 level according to the Common European Framework of Reference (CEFR).[2] All participants were randomly allocated to one of three groups: an input group (n = 32), an input repetition group (n = 30), and a no-input group (i.e., baseline group) (n = 29). All three groups performed one oral task and repeated the same oral task two days later. However, while the input group was exposed to input before performing the immediate oral task, the input repetition group was exposed to input before and after the immediate oral task. The no input group only performed the oral tasks with no input exposure. All learners participated on a voluntary basis.

3.2. Input

The input consisted of three authentic L2 captioned English-language videos about tourist attractions in Iceland and Cairns, taken from the *Viator Travel* YouTube channel (total time = 6 min). The videos were previously used in Duong, Montero Perez, Desmet, and Peters (2021) study and were found to be comprehensible and motivating to Vietnamese L2 learners at A2-B1 level. The lexical profile of the video was checked with Nation's Range software, which showed that the 3000 most frequent word families (without proper nouns) provided 95% coverage and 5000 most frequent word families provided 98% coverage.

3.3. Oral task

Participants were asked to orally describe their own (imaginary) travel experience to Iceland and Cairns in English using pictures as prompts. Participants of the experimental groups (i.e., input and input repetition) were exposed to the input (i.e., the captioned videos). Before watching the videos, they were informed that they would have to perform an oral task on the topic of the videos. Note-taking was not allowed. The instructions aimed to enhance learners' engagement with and attention to the input. In this respect, our task is similar to the ones used in previous studies (e.g., story retelling task in Hoang & Boers, 2016; video summary task in Nguyen & Boers, 2018) in that there is a close relation between the topic of the input and the oral task. Our task differs in that learners need to manipulate and reorganize information processed from the videos to generate their own creative story based on the picture prompts. According to Joe (1998), such a process of active generation can create a cognitive link between new information with existing one, which is particularly valuable for word retention.

Regarding the task prompts, there were seventeen pictures, twelve of which characterize objects shown in the videos (i.e., aboriginal, submarine, glass bottom boat, tropical forest, hot-air balloon, reef, hot spring, sword, armor, northern light, spear, sea

[1] One-hundred and thirty-four learners originally participated in the study. However, data of forty-three participants were excluded: thirty-six participants took note and watched videos with a similar topic by themselves after doing the first speaking tasks; five participants produced narratives of fewer than 100 words; one participant did not complete both vocabulary knowledge tests and one recording has bad sound quality.

[2] Participants' proficiency level was determined based on the entrance test in the format of a TOEIC test which was developed by the local university to test students' English level at the beginning of each semester. The test only contains listening and reading sections. Their total scores ranged from 340 to 750 (M = 501.24, SD = 93.16), which approximately corresponds to the A2 to B1 level according to the CEFR (ETS, n.d.).

walker). We included five additional pictures about objects not visible in the videos but familiar[3] to the learners to enhance their speaking motivation (e.g., horse riding, mountain climbing). These pictures were added because a pilot study had shown that learners were reluctant to speak if they could only describe one or two pictures. The specific instruction of the oral task was as follows:

Imagine that you have just come back to Vietnam from a trip to Iceland and Cairns, and you took seventeen photos of these five places. Retell your travel experience based on these photos. You do not need to talk about all of the photos. You have at most 3 min to prepare and another 3 min to talk. You are not allowed to use a dictionary or any other reference sources.

Following Li, Chen, and Sun's (2015) findings, participants were given 3 min of planning time since this was found to be optimal for learners to prepare what they want to say and how to put these pre-verbal messages into appropriate words and structures. Although planning time may create favorable conditions for lexical mining because learners might take it as an opportunity to rehearse lexis mined from the input, it is difficult to prevent a certain degree of planning irrespective of whether it is formally provided in the treatment or not.

3.4. Measures of lexical mining

Lexical mining was measured by computing the number of shared single words and shared formulaic sequences (FS) between learners' oral texts and the input text. We decided to investigate the mining of FS as well given that learners may struggle to successfully mine FS from input (Hoang & Boers, 2016).

The input text and oral tasks were lemmatized before using the Text Lex Compare function on http://www.lextutor.ca to compute the number of lemmas shared between the oral tasks and the input text. Lemmatizing was important as this technique allowed us to calculate the shared single words on the basis of the base form of the word. Only lemmas of the content words were counted because the function words are high-frequency words and thus are unlikely to be mined from the input. The calculation yielded 310 lemmas in total.

Siyanova-Chanturia and Pellicer-Sánchez (2019) defined FS as conventionalized "strings of letters, words, sounds or other elements" that can be varied in "length, size, frequency, degree of compositionality" but "hold a strong relationship in communicating meaning" (p.5). Thirty-seven FS were selected from the videos. Then, we used SketchEngine (Kilgarriff et al., 2014) to check the sequences' MI (Mutual Information) score in the Open Subtitles corpus (Tiedemann, 2012). FS with an MI score of 3 or higher were included (Hunston, 2002) (see Table 1 for the final list). Two independent researchers, who are EFL Vietnamese university lecturers with more than five years' experience, checked the FS manually. The inter-rater reliability was high (r = .98). We counted successful attempts (i.e., with pronunciation mistakes tolerated) as well as unsuccessful attempts to mine FS. Unsuccessful attempts refer to learners' efforts to mine the FS but the forms are deviant; for example, the order of the FS components was reversed because of the L1 influence (e.g., 'bottom glass boat' for 'glass bottom boat'); one component was replaced by another word in the FS (e.g., 'southern light' for 'northern light'); one component was missed for three-word FS (e.g., air balloon, glass boat).

3.5. Prior vocabulary knowledge tests

A receptive vocabulary size test and a productive levels test were used to measure the participants' prior vocabulary knowledge. Nguyen and Nation's (2011) English-Vietnamese version of the Vocabulary Size Test (VST) was adopted to measure the learners' receptive vocabulary knowledge. The VST is a frequency-based meaning recognition test which samples 10 words from 14 frequency bands of 1000 words (1K–14K). Previous research showed that the test procedure was too long and a short version (70-item) of the original bilingual test (140-item) was proved valid and reliable (e.g., Duong et al., 2021). Hence, we used the 70-item version in the current study. A good internal consistency was found (Cronbach's alpha = .87, n = 91).

Learners' productive vocabulary knowledge was measured by Laufer and Nation's (1999) Productive Levels Test. This test aims to estimate the learners' controlled productive word knowledge (i.e., words that a learner can produce only when prompted) at the 2000, 3000, 5000 and 10,000-frequency levels, and academic words from the University Word List. The test has a written gap-filling format with the first letter of each target word given as prompts. A good internal consistency was found (Cronbach's alpha = .94, n = 91).

3.6. Working memory tests

Separate tests were used to assess different components of WM. Specifically, we used a forward digit-span task to measure phonological short-term memory ability and a backward digit-span task to measure learners' ability to manipulate phonological information. Nonverbal tasks were considered appropriate as a close link between the learners' ability to repeat nonwords and to learn the phonological forms of words was found (Gathercole, 2006). We also administered an operation-span task (Ospan) (Turner & Engle, 1989) to measure the executive and attention-regulatory functions of WM following the evidence that the executive system in WM is domain-general in nature rather than specific to language (Baddeley, 2003).

For the digit-span tasks (both forward- and backward-span), participants listened to a pre-recorded sequence of digits. In the forward-span task, participants were asked to repeat the digits in the presented order immediately after listening whereas they had to repeat the digits in the reverse order in the backward-span task. The forward-digit span task consisted of 8 spans, going from 3 to 9

[3] Learners have learned vocabulary related to these pictures in their current learning program.

Table 1
List of formulaic sequences.

	Item	FoE	Frequency	MI score	Congruency
1	Tropical climate	1	132	7.6	1
2	Getaway spot	1	25	3.45	0
3	Check out	2	6660	6.74	0
4	Go up	1	28,001	8.86	0
5	Go away	1	3418	8.35	0
6	Hot-air balloon	1	87	8.98	0
7	Rain forest	1	592	7.5	0
8	Hop in	1	126	5.31	1
9	Get wet	1	1892	3.64	0
10	Unique perspective	1	176	6.07	0
11	Spear fishing	1	16	3.56	0
12	Make sure	1	22,910	11.53	0
13	Volcanic basalt	1	3	5.24	0
14	Faux fur	1	5	4.37	0
15	Bear in mind	1	176	7.33	0
16	Northern light	1	217	4.82	0
17	Pay off	1	3832	8.01	0
18	Geothermal plant	2	60	4.09	0
19	Hot spring	1	416	6.52	0
20	Do something right	1	2511	4.19	0
21	Feel like	1	807	7	1
22	Cuddle a koala	1	3	7.34	0
23	Reach up to	1	682	3.69	0
24	Come out	1	2182	6.86	0
25	Take a minute	1	5609	5.83	0
26	Above the sea level	1	1732	5.77	1
27	Watch the sunrise	1	132	4.38	1
28	Walk in a line	1	132	5.61	1
29	Plan a trip	1	355	4.23	1
30	Get started	1	389	4.62	0
31	Laid-back attitude	1	13	4.27	0
32	Glass bottom boat	1	13	4.64	1

Note. FoE: Frequency of encounters.

digits with 2 sets per span. The backward-digit span task consisted of 8 spans, going from 2 to 8 digits with 2 sets per span. The participants needed to accurately recall the digits of at least one set to continue to the next span. The participants' score (max = 16) equals the total number of correctly recalled sets.

A Vietnamese version of the Ospan task was adapted on the Dutch version [4](De Neys, d'Ydewalle, Schaeken, & Vos, 2002), using Affect 5.0 software (https://ppw.kuleuven.be/apps/clep/affect5/repository.php). In the task, participants were required to solve a series of math operations and try to remember the word that appeared after each operation. The test had 15 sets with 2–6 operation-word strings per set. The operation-word strings were presented on PC. After each complete set, the participants were asked to write down the words in the presented order on an answer sheet. The participants' span score (max = 60) equals the sum of words correctly recalled in the exact order.

3.7. *Procedure*

Data of all participants were collected in three one-to-one and face-to-face sessions over a 1-week period with the third co-author, an EFL teacher in Vietnam. The first author was also present via Skype during the data collection to check treatment fidelity.[5] During the first session, the input group and the input repetition group watched the videos twice on a computer screen before performing an immediate oral task. The participants knew that the topic of the immediate oral task would be similar to the topic of the videos. After completing the output task, only the input repetition group watched the videos once more. They were asked to watch the videos for comprehension and personal opinion (i.e., watch the videos again and tell the teacher which destination is their favorite one). Further, our participants were not encouraged to take notes, a strategy commonly used in listening activities, when watching videos (as in Nguyen & Boers, 2018) or to use the notes during speaking (as in Hoang & Boers, 2016), to avoid a confounding effect of learners' note-taking skills (e.g., learners mine fewer words in the oral performance because of poor note-taking rather than a lack of input engagement). The no-input group only performed the oral task without watching the videos. After the treatment, all participants had a 15-min break before doing the forward and the backward-digit span tasks (15 min for both tasks). During the second session, which

[4] We adopted exactly the same set of 66 operations from the Dutch version. However, we replaced high frequency Dutch nouns by high frequency Vietnamese nouns, randomly selected from the Corpora of Vietnamese Texts (Pham, Kohnert, & Carney, 2008).

[5] The first author took notes during the treatment but did not give further support to the third author.

took place two days after the first one, all participants performed the same surprise oral task without watching the videos. They had a 10-min break before doing the Ospan task (30 min). The receptive and productive vocabulary knowledge tests were completed during the third session, with a 15-min break in between.

The participants' oral performances and their responses in the forward and backward-digit span tests were recorded by a handheld audio recorder. Half of the recordings were transcribed by the first author and half by the third author. The transcriptions were then cross-verified.

3.8. *Statistical analysis*

To explore the effects of input, input repetition, and individual differences on lexical mining in the immediate and the repeal oral tasks, a repeated measures MANCOVA analysis in SPSS 27 was performed. All participants completed an immediate oral task and a repeat oral task; thus, *task time* (immediate vs. repeat) was a within-subjects factor. *Input exposure* (input repetition vs. input vs. no input) was a between-subjects factor. The dependent variables were shared lemmas and shared FS. Covariates were scores of the working memory tests and the prior vocabulary knowledge tests. Since all covariates displayed Pearson correlation below 0.70, the threshold for multicollinearity effect to occur (e.g., Crossley, Salsbury, McNamara, & Jarvis, 2011), they were all included in the regression model. Assumptions including outlier bias, normality, and homogeneity of variance-covariance were approximately met.[6] An alpha level of 0.05 in Pillai Trace's test[7] was taken as the level of statistical significance. Then, follow-up univariate tests were performed to answer specific research questions with the adjusted p-value (see more in Results section).

4. Results

4.1. *Participants characteristics*

4.1.1. *Prior vocabulary knowledge*

Table 2 shows that participants in the three groups performed similarly on the receptive test and the productive vocabulary tests (for detailed information per test level, see Appendix A). An ANOVA showed that the three groups did not significantly differ in their receptive ($F(2, 87) = .195$, $p = .823$) and productive vocabulary knowledge ($F(2, 87) = 2.784$, $p = .067$).

4.1.2. *Working memory*

Table 3 shows that the participants performed well in forward-span, backward span tests and Ospan task. An ANOVA did not show a significant difference between the three groups in terms of working memory: forward-span ($F (2, 89) = .822$, $p = .443$), backward-span ($F(2, 89) = .249$, $p = .780$), and Ospan task ($F(2, 89) = 1.43$, $p = .243$).

4.2. *Research question 1: lexical mining in the immediate and the repeat oral tasks*

Table 4 shows that the number of shared lemmas and shared FS between the input text and the immediate oral task was higher for the input and input repetition groups than the no input group. Overall, participants who were exposed to input could mine approximately 10% of the total lemmas and 13% of the target FS in the immediate oral task. It should be noticed that the no-input group used 7% of the words occurring in the input, suggesting that these were known words. This means that the actual number of words mined may be more limited. Yet, the no-input group did not produce any of the target FS, indicating that these FS might have been novel to the participants. Regarding the difference between lexical mining in the immediate and the repeat oral task, Table 4 shows that the learners mined almost the same number of lemmas in the repeat task as in the immediate task. However, the number of mined FS was lower in the repeat task than in the immediate task.

Table 5 reports the successful and unsuccessful attempts of mined FS. The findings showed that around 80% of attempts to mine FS were successful. As for the total number of tokens and lemmas the participants produced in the immediate and the repeat oral tasks, see Appendix B.

The repeated-measures MANCOVA revealed a significant effect of input exposure on the overall lexical mining ($F (4, 164) = 11.159$, $p < .001$, $\eta 2 = .214$, power $= 1.000$) with a large effect size.[8] The effect of task time was not significant ($F (2,81) = .293$, $p = .746$). However, there was an interaction effect between input exposure and task time on the overall lexical mining ($F(4, 164) = 3.795$,

[6] One outlier was detected using Mahalanobis Distances and was removed from the dataset. Thus, the final dataset contained data of 90 participants. The Shapiro-Wilk's tests indicated a violation of normality of the number of shared FS. Log-transformation did not improve the normality. Yet, the P–P and Q-Q plots showed that the distribution of residuals of the number of shared FS was only slightly skewed. The Box's M test indicated that the homogeneity of variance-covariance was violated. As the Box's M test is extremely sensitive to departures from normality, the Levene's test was checked for the homodestacity per dependent variable. It was shown that the homodestacity assumption was met for all variables except for the shared FS in the immediate oral task ($p < .001$). As MANCOVA analyses were robust to normality assumption (Norman, 2010), this method was used to run the data analysis.

[7] Pillai Trace's test was used instead of Wilk's Lambda as Pillai Trace's test was considered robust to the violation of homogeneity of covariance-variance (Olson, 2012).

[8] The values of $\eta 2$ were interpreted as follows: $\eta 2 > 0.0099$ = small, $\eta 2 > 0.0588$ = moderate, and $\eta 2 > 0.1379$ = large (Cohen, 1992).

Table 2
Descriptive statistics for the receptive and productive vocabulary knowledge tests.

	Receptive test (Max = 70)		Productive test (Max = 90)	
	Mean (SD)	95% CI	Mean (SD)	95% CI
Input repetition (n = 32)	41.68 (10.0)	23.00–61.00	28.18 (15.33)	15.00–59.00
Input (n = 29)	43.36 (7.93)	23.00–56.00	31.57 (12.01)	22.00–55.00
No input (n = 29)	43.95 (7.79)	30.00–59.00	35.59 (8.55)	18.00–50.00

Table 3
Mean scores and standard deviations (in brackets) for the three working memory tests.

	Forward-span (max = 16) Mean (SD)	Backward-span (max = 16) Mean (SD)	OSPAN (max = 60)> Mean (SD)
Input repetition (n = 29)	14.43 (2.09)	14.06 (2.50)	50.86 (8.43)
Input (n = 32)	13.50 (3.45)	13.59 (2.78)	51.18 (11.13)
No input (n = 29)	14.10 (2.90)	13.96 (3.05)	46.58 (14.76)

Table 4
Descriptive statistics of the shared lemmas and shared FS.

	Immediate task		Repeat task	
Shared lemmas	Mean (SD)	95% CI	Mean (SD)	95% CI
Input repetition (n = 29)	32.90 (9.27)	14–54	33.00 (10.22)	19–71
Input (n = 32)	33.96 (10.18)	18–56	33.46 (11.16)	15–60
No input (n = 29)	23.03 (5.68)	11–37	22.20 (6.14)	12–38
Shared FS	Mean (SD)	95% CI	Mean (SD)	95% CI
Input repetition (n = 29)	3.90 (2.32)	0.00–6.00	2.86 (1.69)	1.00–8.00
Input (n = 32)	4.37 (2.72)	0.00–10.00	2.43 (1.96)	0.00–8.00
No input (n = 29)	0.68 (0.89)	0.00–3.00	0.68 (0.80)	0.00–3.00

Note. Total number of lemmas = 310; total number of FS = 32.

Table 5
Means and standard deviations of successful and unsuccessful attempts to mine FS.

	Immediate task			Repeat task		
	Successful Mean(SD)	Unsuccessful Mean (SD)	Total Mean(SD)	Successful Mean (SD)	Unsuccessful Mean (SD)	Total Mean (SD)
Input repetition (n = 29)	2.68 (1.75)	1.10 (1.08)	3.90 (2.32)	1.86 (1.22)	1.00 (0.94)	2.86 (1.69)
Input (n = 32)	3.53 (2.72)	0.84 (0.98)	4.37 (2.72)	1.84 (1.52)	0.59 (0.94)	2.43 (1.96)
No input (n = 29	0.51 (0.87)	0.17 (0.38)	0.68 (0.89)	0.41 (0.62)	0.27 (0.70)	0.68 (0.80)

p = .006, $\eta 2$= .085, power = .885). Learners' receptive vocabulary knowledge score was the only variable which significantly predicted overall lexical mining (F (2, 81) = 4.529, p = .014, $\eta 2$ = .101, power = .757). Follow-up univariate analyses were conducted to answer each research question with adjusted p-values for multiple comparisons (p = .05/4 = .0125).

Concerning lexical mining in the immediate oral task, Table 6 shows a significant effect of input exposure on the number of shared lemmas (F (2, 82) = 13.606, p < .001) and shared FS (F (2, 82) = 22.864, p < .001). Post hoc pairwise comparisons displayed no significant difference between the input and the input repetition groups in the number of shared lemmas (p = .828) and shared FS (p = .650), which was expectable as these two groups had the same number of input encounters before completing the immediate oral task. A significant difference was found only between the input and the no input group (p < .001 for shared lemmas, p < .001 for shared FS) as well as between the input repetition and the no input group (p = .001 for shared lemmas, p < .001 for shared FS).

To check if the difference in lexical mining between the immediate and the repeat oral task was significant, repeated-measures ANCOVAs were computed. The analyses showed that the effect of task time was non-significant for the shared lemmas (F(1,82) = .009, p = .926) but significant for the shared FS (F (1,82) = 23.35, p < .001). This finding indicates that the learners indeed mined significantly fewer FS in the repeat oral task.

4.3. Research question 2: The effect of input repetition on lexical mining in the repeat oral task

Table 4 shows that the input repetition and the input groups could mine more lemmas and FS than the no input group. However, the number of lemmas and FS mined was almost the same between the input repetition and the input groups. Table 7 reveals a significant effect of input exposure on the shared lemmas (F(2, 82) = 15.625, p < .001) and the shared FS in the repeat oral task (F(2, 82) = 13.636, p < .001). Post hoc pairwise comparisons displayed no significant difference between the input and the input repetition groups

Table 6
One-way ANCOVAs of the shared lemmas and shared FS in the immediate oral task.

	DVs	*df*	*F*	*p*	*η*2	Power
Input exposure	Shared lemmas	2	13.606	<.0001	.249	.998
	Shared FS	2	22.864	<.0001	.358	1.000
Receptive	Shared lemmas	1	4.064	.047	.047	.513
	Shared FS	1	1.607	.209	.019	.240
Productive	Shared lemmas	1	6.919	.010	.078	.739
	Shared FS	1	1.035	.312	.012	.171
Forward-span	Shared lemmas	1	.586	.446	.007	.118
	Shared FS	1	.000	.989	.000	.050
Backward-span	Shared lemmas	1	2.563	.113	.030	.353
	Shared FS	1	.784	.379	.009	.141
Ospan	Shared lemmas	1	2.640	.108	.031	.362
	Shared FS	1	1.495	.225	.018	.227
Error	Shared lemmas	82				
	Shared FS	82				

Table 7
One-way ANCOVAs of the shared lemmas and the shared FS in the repeat oral task.

	DV	*df*	*F*	*p*	*η*2	Power
Input exposure	Shared lemmas	2	15.625	<.001	.276	.999
	Shared FS	2	13.636	<.001	.250	.998
Receptive	Shared lemmas	1	11.396	.001	.122	.916
	Shared FS	1	2.732	.102	.032	.372
Productive	Shared lemmas	1	.415	.521	.005	.098
	Shared FS	1	1.188	.279	.014	.190
Forward-span	Shared lemmas	1	.128	.722	.002	.064
	Shared FS	1	.539	.465	.007	.112
Backward-span	Shared lemmas	1	1.741	.191	.021	.257
	Shared FS	1	.375	.542	.005	.093
Ospan	Shared lemmas	1	1.758	.189	.021	.258
	Shared FS	1	1.091	.299	.013	.178
Error	Shared lemmas	82				
	Shared FS	82				

in the number of shared lemmas ($p = 1.000$) and shared FS ($p = .792$). A significant difference was found only between the input and the no input group ($p < .001$ for shared lemmas, $p < .001$ for shared FS) as well as between the input repetition and the no input group ($p < .001$ for shared lemmas, $p < .001$ for shared FS). These findings indicate that input repetition did not have an added effect on participants' mining of lemmas and FS in the repeat oral task.

To investigate on which lexical mining measure the interaction effect occurs, separate repeated-measures ANCOVAs were performed. A significant interaction effect between *input exposure* and *task time* was found on the shared FS ($F(2, 82) = 13.310$, $p = .001$, $\eta 2 = .152$, power = .932) but not on the shared lemmas ($p = .858$). The interaction plot (see Fig. 1) shows that the effect of task repetition on FS mining might differ when the input is repeated.

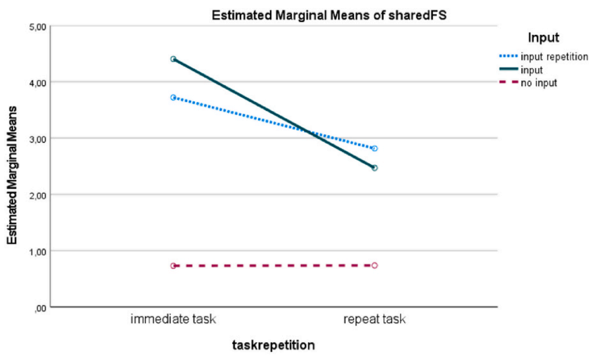

Fig. 1. Interaction effect of input exposure and task time on shared FS.

4.4. Research question 3: The moderating effects of individual differences

In the immediate oral task, Table 6 shows a significant, albeit weak relationship between participants' productive vocabulary knowledge ($F(1, 82) = 6.919$, $p = .010$, $\eta 2 = .078$, power $= .739$) and the number of shared lemmas. Learners' prior vocabulary knowledge, however, did not play a role in the mining of FS. In the repeat oral task, it was only receptive vocabulary knowledge which significantly mediated participants' mining of lemmas ($F(1, 82) = 11.396$, $p = .001$, $\eta 2 = .122$, power $= .916$) (see Table 7). Working memory did not play a role in participants' lexical mining, neither in the immediate nor in the repeat oral tasks.

5. Discussion

5.1. Lexical mining in an immediate and a repeat oral task

The present study confirms Hoang and Boers' (2016) finding that L2 learners can mine single words from audiovisual input in their output tasks. However, the proportion of words successfully mined was lower in our study than in Hoang and Boers's (2016) study (10% vs. 68%). We assume that the difference in findings between the two studies might be attributed to differences in learning conditions. In the present study, note-taking was not allowed and the input text was not at learners' disposal while speaking. In addition, lexical items were operationalized differently (i.e., lemmas) and participants' proficiency level was lower (i.e., A2-B1 level) in this study than in Hoang and Boer's study (i.e., B1–B2 level).

Our findings also mirror those of Hoang and Boers by showing that L2 learners can mine FS but the number of mined FS was minimal compared to lemmas. Taken together, it seems that it is more challenging to mine FS than single words. In the current study, we assume that learners' lexical mining might also have been influenced by the prompts in the oral task. The lemmas related to these prompts may have been more concrete than the FS, which might explain why more lemmas were mined than FS. Also, the difference in the retention of mined lemmas and FS in the repeat oral task indicates that the task repetition effect might be influenced by the type of lexical item.

The current study adds more insights to the mining of FS by taking into account the successful as well as unsuccessful attempts of FS mining. The fact that most of the attempts were successful (80%) seems a promising finding, especially since learners watched the videos only twice before speaking. Yet, there stands a chance that the participants may have had partial knowledge of some FS. A qualitative analysis of the participants' spoken output shows that FS successfully mined often contained word components that were frequent and concrete (e.g., rain forest, above the sea level), which might have facilitated learners' mining of FS.

5.2. The effect of input repetition on lexical mining in a repeat oral task

The results indicate that reviewing the videos after the immediate oral task does not help the mining of lemmas and FS in the repeat task. This finding suggests that the input-output-input cycle might only be effective in triggering learners' need to reflect on and notice the form-meaning mapping of new words (as shown in Nguyen & Boers, 2018) and not in lexical mining. The non-significant effect of the repeated input might also be explained in light of Laufer and Hulstijn's (2001) involvement load hypothesis which emphasizes the importance of three components – need, search, and evaluation in any learning task for improving vocabulary uptake. These components seemed to be absent in the video reviewing activity which only drew learners' attention to the video content and did not explicitly stimulate learners to reflect on words mined from the videos. Also, it cannot be fully excluded that some participants might not have taken the repeated input seriously due to fatigue or boredom, which could have affected their attention. However, it was found that input repetition seems to slow down the attrition process of mined FS. We assume that this is likely due to the perceived novelty of these FS compared to the single words that make them relatively salient for the participants and thus they remembered the mined FS longer.

5.3. The role of individual differences

Our results revealed that the predictive power of learners' prior vocabulary knowledge seemed to be stronger in the repeat task than in the immediate one. In the immediate task, input exposure might have obscured the effect of learners' prior vocabulary knowledge. Furthermore, the strong association between receptive vocabulary knowledge and the number of lemmas mined in the repeat task might result from the fact there was more variance in the receptive than the productive vocabulary knowledge scores.

In terms of WM, no significant moderating effect was found. One possible reason is that the participants' scores were very similar as evidenced by the small variance in WM score distributions in all three tests. Further, a ceiling effect observed in the forward-span and backward-span tests (i.e., mean scores are relatively close to the maximum obtainable scores) might have suppressed the correlation between WM and mined lemmas/FS. Therefore, future studies may use WM tests with longer sequences of digits to increase variance among learners.

In addition, a few studies have suggested that the effect of phonological WM may be more important amongst beginning L2 learners (Juff & Harrington, 2011; Mitchell, Jarvis, O'Malley, & Konstantinova, 2015; O'Brien, Segalowitz, Collentine, & Freed, 2006). Perhaps this is also the reason why the predictive role of the forward digit span scores was very limited in our study. Also, the sample size, which is smaller in the present study than in previous studies (e.g., Kormos & Sáfár, 2008), might have reduced the likelihood of finding correlations. Another reason might be that working memory may be more predictive of vocabulary learning (e.g., French, 2004; Gathercole, Willis, Ellis, & Baddeley, 1992; Montero Perez, 2020) rather than vocabulary use (e.g., lexical mining in our study), as similarly argued by O'Brien et al. (2006).

6. Conclusion

Overall, the findings of the present study suggest that L2 learners can mine words and FS in an oral task immediately after watching videos as well as two days later. However, watching the videos again after the immediate oral task did not help learners mine more words and FS in the repeat oral task. It was only effective in slowing down the loss of mined FS. The study also highlighted the significant role of prior vocabulary knowledge for single word mining.

This study also has some limitations. First, the participants are university students at A2-B1 proficiency level, making it difficult to generalize the findings to participants of other profiles. Thus, future studies could explore L2 learners' lexical mining across levels of English proficiency. Second, the interval between the immediate and the repeat oral tasks was two days, which may have been relatively short and might not represent the situations when learners have to repeat the task after longer intervals. It has been shown that a choice of interval between task performances can influence the enhancement of different aspects of oral performance (e.g., complexity, accuracy, fluency) (Bui, Ahmadian, & Hunter, 2019). Yet no research has explored the extent to which task interval influences lexical mining in repeat performances, which warrants future studies. Finally, our findings were solely based on one type of oral task (i.e., picture-prompted narrative task). Future studies thus might need to explore whether lexical mining will also occur in other task types (e.g., narrative tasks without prompts, conversational tasks).

The study has a number of pedagogical implications. For instance, teachers should take into account the differences between the uptake of single words and FS in setting up their vocabulary teaching programs, as also suggested by Alali and Schmitt (2012). In addition, as the kind of input repetition proposed in this study did not yield a strong effect, teachers are advised to try out other pedagogical activities such as those recommended by Lynch (2018) (e.g., providing feedback, correcting the transcript of the immediate oral task) or employ a strategy to increase learners' cognitive involvement with the repeated input (e.g., implicitly by informing learners of the repeat task, or explicitly by asking learners to watch and compare their use of words in the narrative with the use of words in the repeated input).

CRediT authorship contribution statement

P.T. Duong: Conceptualization, Methodology, Data collection, Formal analysis, Writing – original draft. **M. Montero Perez:** Conceptualization, Writing – review & editing. **L.Q. Nguyen:** Data collection, Writing – review & editing. **P. Desmet:** Conceptualization, Writing – review & editing. **E. Peters:** Conceptualization, Methodology, Writing – review & editing.

Appendix A

TABLE 1
Descriptive statistics for the productive levels test

	2K (Max = 18) 95% CI	3K (Max = 18) 95% CI	5K (Max = 18) 95% CI	10K (Max = 18)95% CI	UWL (Max = 18)95% CI	Total (Max = 90) 95% CI
Input repetition (n = 32)	12.07 (3.99)	5.59 (3.90)	4.37 (3.24)	1.62 (2.45)	4.37 (3.83)	28.18 (15.33)
	4.00–17.00	0.00–13.00	0.00–14.00	0.00–8.00	0.00–14.00	15.00–59.00
Input (n = 29)	14.11 (3.02)	9.61 (3.46)	7.65 (2.52)	2.73 (2.14)	7.46 (3.21)	31.57 (12.01)
	7.00–18.00	3.00–14.00	3.00–12.00	0.00–7.00	0.00–12.00	22.00–55.00
No input (n = 29)	13.96 (2.27)	8.55 (2.75)	5.00 (2.41)	2.14 (1.53)	5.92 (3.11)	35.59 (8.55)
	10.00–18.00	3.00–14.00	0.00–10.00	0.00–5.00	0.00–12.00	18.00–50.00

Note. UWL: University Word Level.

Appendix B

TABLE 2
Means and standard deviations of total tokens and lemmas produced

	Tokens		Lemmas	
	Immediate task Mean (SD)	Repeat task Mean (SD)	Immediate task Mean (SD)	Repeat task Mean (SD)
Input repetition (n = 29)	229.66 (57.34)	228.70(54.46)	71.70 (19.71)	75.53 (20.19)
Input (n = 32)	235.68 (83.61)	232.56 (67.89)	74.43 (25.54)	76.68 (22.03)
No input (n = 29)	242.62 (66.25)	250.72 (63.46)	75.41 (21.88)	74.86 (18.54)

References

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal, 3*(2), 153–180. https://doi.org/10.1002/tesj.13

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Boston, J. S. (2008). Learner mining of pre-task and task input. *ELT Journal, 62*(1), 66–76. https://doi.org/10.1093/elt/ccm079

Bui, G., Ahmandian, M. J., & Hunter, A.-M. (2019). Spacing effects on repeated L2 task performance. *System, 81*, 1–13. https://doi.org/10.1016/j.system.2018.12.006

Bygate, M. (2018). Introduction. In M. Bygate (Ed.), *Learning Language through task repetition* (pp. 1–25). Amsterdam: John Benjamins.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing, 28*(4), 561–580. https://doi.org/10.1177/0265532210378031

De Neys, W., d'Ydewalle, G., Schaeken, W., & Vos, G. (2002). A Dutch, computerized, and group administrable adaptation of the operation span test. *Psychologica Belgica, 42*, 177–190. https://doi.org/10.5334/pb.993

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34*(1), 5–34. https://doi.org/10.1017/S0272263111000489

ETS (n.d.). Mapping the TOEIC tests on the Common European Framework of reference for languages. https://www.ets.org/s/toeic/pdf/toeic-cefr-flyer.pdf.

Duong, P. T., Montero Perez, M., Desmet, P., & Peters, E. (2021). Learning vocabulary in spoken input- and output-based tasks. *TASK, 1*(1), 100–126.

French, L. (2004). *Phonological working memory and L2 acquisition: A developmental study of Francophone children learning English in Quebec*. Edwin Mellen Press.

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research, 20*, 321–340. https://doi.org/10.1177/1362168815570142

Gathercole, S. E., Willis, C. S., Ellis, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513–543. https://doi.org/10.1017/S0142716406060383

Hoang, H., & Boers, F. (2016). Re-telling a story in a second language: How well do adult learners mine an input text for multiword expressions? *Studies in Second Language Learning and Teaching, 6*(3), 513–535. https://doi.org/10.14746/ssllt.2016.6.3.7

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *Modern Language Journal, 104*(3), 550–566. https://doi.org/10.1111/modl.12661

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*, 357–377.

Juff, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Kim, Y., Crossley, S., Yung, Y., Kyle, K., & Kang, S. (2018). The effects of task repetition and task complexity on L2 lexical use. In M. Bygate (Ed.), *Learning language though task repetitions* (pp. 75–96). Amsterdam: John Benjamins.

Kilgarriff, A., Baisa, V., Bušta, J., et al. (2014). The sketch engine: Ten years on. *Lexicography ASIALEX, 1*, 7–36. https://doi.org/10.1007/s40607-014-0009-9

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33–51. https://doi.org/10.1177/026553229901600103

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1–26. https://doi.org/10.1093/applin/22.1.1

Li, L., Chen, J., & Sun, L. (2015). The effects of different lengths of pre-task planning time on L2 learners' oral test performance. *TESOL Quarterly, 49*, 38–66. https://doi.org/10.1002/tesq.159

Lynch, T. (2018). Perform, reflect, recycle: Enhancing task repetition in second language speaking classes. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 193–222). Amsterdam: John Benjamins.

Lynch, T., & Maclean, J. (2000). A case of exercising: Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 141–162). Harlow: Pearson Education.

Mitchell, A. E., Jarvis, S., O'Malley, M., & Konstantinova, I. (2015). Working memory measures and L2 proficiency. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 270–284). Bristol: Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781783093595-019.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *EUROSLA Monographs series: 2. L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (pp. 57–78).

Norman, G. (2010). Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Sciences Education, 15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Noreillie, A.-S., Desmet, P., & Peters, E. (2020). Factors predicting low-intermediate French learners' vocabulary use in speaking tasks. *Canadian Modern Language Review, 76*(3), 194–217. https://doi.org/10.3138/cmlr-2019-0018

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition, 42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Nguyen, C. D., & Boers, F. (2018). The effect of content retelling on vocabulary uptake from a TED Talk. *TESOL Quarterly, 53*(1), 5–29. https://doi.org/10.1002/tesq.441

Nguyen, L. T. C., & Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42*(1), 86–99. https://doi.org/10.1177/0033688210390264

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistic, 27*, 377–402. https://doi.org/10.1017/S0142716406060322

Olson, C. L. (2012). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69*(348), 894–908. https://doi.org/10.2307/2286159

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

Peters, E. (2020). Factors affecting the learning of single-word items. In S. Webb (Ed.), *The routledge Handbook of vocabulary studies* (pp. 125–142). Routledge.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through watching a single episode of L2 television. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Pham, G., Kohnert, K., & Carney, E. (2008). Corpora of Vietnamese texts: Lexical effects of intended audience and publication place. *Behavior Research Methods, 40*, 154–163. https://doi.org/10.3758/BRM.40.1.154

Read, J. (2020). Key issues in measuring vocabulary knowledge. In S. Webb (Ed.), *The routledge Handbook of vocabulary studies* (pp. 125–142). Routledge.

Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogical tasks: Second language learning, teaching, and assessment* (pp. 119–140). London: Longman.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. London, UK: Routledge.

Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. *Proceedings of the 8th international conference on language resources and evaluation*. Istanbul: LREC.

Turner, M. L., & Engle, R. W. (1989). Is working-capacity task-dependent? *Journal of Memory and Language, 28*, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research, 24*(4), 540–556. https://doi.org/10.1177/1362168818799371

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal, 47*(1), 64–75. https://doi.org/10.1080/09571736.2016.1191527

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

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System, 70*, 38–49. https://doi.org/10.1016/j.system.2017.09.012