Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 13 (1). 2023. 101-124
https://doi.org/10.14746/ssllt.29727
http://pressto.amu.edu.pl/index.php/ssllt

The impact of input, input repetition, and task repetition on L2 lexical use and fluency in speaking

Phuong-Thao Duong ✉
KU Leuven, Belgium
http://orcid.org/0000-0002-0900-9357
duongthao204@gmail.com

Maribel Montero Perez
Ghent University, Belgium
https://orcid.org/0000-0002-0868-588X
maribel.monteroperez@ugent.be

Long Quoc Nguyen
FPT University, Vietnam
https://orcid.org/0000-0002-4674-7199
quocnl2@fe.edu.vn

Piet Desmet
KU Leuven & imec, Belgium
https://orcid.org/0000-0002-9849-0874
piet.desmet@kuleuven.be

Elke Peters
KU Leuven, Belgium
https://orcid.org/0000-0001-7273-3850
elke.peters@kuleuven.be

Abstract
The present study investigates the impact of meaningful input on L2 learners' vocabulary use and their fluency in oral performance (immediate and repeat tasks), as well as whether the effects are mediated by learners' prior vocabulary

knowledge and working memory. Ninety university students learning English as a foreign language were randomly assigned to one of three groups: input (*N* = 29), input repetition (*N* = 32), and no-input (i.e., baseline group) (*N* = 29). The input group watched L2 videos prior to performing an immediate oral task, whereas the input repetition group watched the same videos not only before but also after the immediate oral task. The no-input group only performed the oral tasks without watching the videos. The three groups repeated the same oral task after two days. Results did not show a significant effect of task repetition, input, and input repetition on learners' lexical use and fluency. However, the fluency and lexical complexity in learners' L2 speech can be predicted by their receptive vocabulary knowledge and working memory capacity to some extent.

Keywords: fluency; input; input repetition; lexical use; speaking; task repetition

1. Introduction

There is general consensus that input is critical for second or foreign language (L2) learning, including vocabulary learning (Webb, 2020). Research shows that learners can learn new words while reading, listening, reading-while-listening and watching TV (for a review, see Webb, 2020); however, the focus has been mainly on the form-meaning link. Very few studies have looked at how L2 learners use vocabulary after being exposed to input. Pedagogically, this is surprising as the goal of any vocabulary program is not only to increase learners' vocabulary knowledge, but also to put learners' vocabulary knowledge into communicative use (Laufer & Nation, 1995). Research has shown that learners tend to reuse single words and multiword units from input in the follow-up oral task (e.g., Duong et al., 2021b; Hoang & Boers, 2016; Nguyen & Boers, 2018). However, how input influences learners' ability to process words from their existing lexical repertoire remains unclear. To fill the gap, the present study aims to investigate the effect of input (i.e., watching L2 captioned videos) and input repetition (i.e., watching the same videos before and after performing an oral task) on L2 learners' lexical use in immediate and repeat oral tasks. We decided to measure oral fluency as well given the established evidence on the trade-off between lexical complexity and fluency in L2 oral performance (Skehan, 2009). Furthermore, this study also explored whether working memory and prior vocabulary knowledge influence the impact of input, input repetition, and task repetition on L2 learners' lexical use and fluency as these two factors have been shown to play a role in L2 oral development (Duong & Le, 2022; Duong et al., 2021b; O'Brien et al., 2006).

2. Literature review

2.1. The role of input for L2 lexical use

Recent research has demonstrated that learners can reuse/mine words from meaningful input, such as pre-task recordings (Boston, 2008) and L2 videos (e.g., Duong et al., 2021b; Hoang & Boers, 2016) in their follow-up oral performance as long as the input is relevant to the task. Yet, it remains unclear whether input exposure has an effect on aspects of lexical use, such as lexical sophistication and lexical diversity. In light of Levelt's (1989) speech processing model, Bygate (1996) and Willis (1996) proposed that exposing learners to relevant input could help reduce learners' cognitive load during task performance. In particular, they argue that being aware in advance of what they will talk or write about through input exposure can help learners shortcut the time needed for building the speech content from scratch and have more time to formulate the language needed to express their ideas. Therefore, we assume that input exposure might improve learners' speech in terms of lexical use.

2.2. The impact of task repetition on L2 lexical use and fluency

Research shows that task repetition can facilitate lexical use, with task repetition defined as the repetition of either "the same or slightly altered tasks – whether whole tasks or parts of a task" (Bygate & Samuda, 2005, p. 43). However, the findings are mixed. For instance, it was pointed out that university students who learned Spanish as L2 could use more low-frequency words (i.e., beyond K3-level) and more diverse words (e.g., Gass et al., 1999) after repeating the same narrative tasks twice. Fukuta (2016) conducted a study on 28 upper-intermediate Japanese learners of English and found that their word choice became more diverse and more accurate in the repeat oral task. Wang (2014) used the same narrative task as Gass et al. (1999) but did not find an effect of task repetition on lexical diversity. This might have been caused by the discrepancy in the learners' L2 levels (advanced vs. intermediate) in these two studies.

It has been widely found that task repetition is useful for fluency development (for a review, see Bygate, 2018). The evidence, however, has been established mainly for generic measures of fluency (e.g., speech rate, frequency/length of pauses). Lately, more attention has been paid to finer-grained measures of fluency (e.g., location of pauses). An important study has been conducted by Lambert et al. (2017) on A2-C1 Japanese learners of English. They found the improvement of speech rate and the declination of between-clause pausing right from the second repetition while within-clause pausing and frequency of repairs only declined from the fourth repetition. This finding indicates that the effect of task repetition can vary depending on the measures and the amount of repetition. However, as the number of studies that have looked at both generic measures and pause location remains scarce, more research is warranted.

In particular, an increasing number of studies have been looking into whether task repetition is beneficial for both fluency and lexical use development; however, the findings are mixed. Khatib and Farahanynia (2020) reported that task repetition promoted L2 learners' fluency but not their lexical use (i.e., lexical sophistication or diversity). In contrast, Fukuta (2016) showed that task repetition had a positive effect on lexical diversity rather than the overall fluency. Duong and Le (2022) surprisingly did not find an effect of task repetition on either lexical use (lexical diversity and words beyond K3+ level) or fluency.

So far, task repetition has often been used in conjunction with other techniques (e.g., pre-task planning, online planning, corrective feedback, post-task transcribing) to foster the concurrent development of accuracy, fluency, and complexity (e.g., for a review see Ellis et al., 2020). To the best of our knowledge, no study has looked into whether task repetition combined with meaningful input exposure can aid the development of lexical use and fluency.

2.3. The potential of input repetition

Nguyen and Boers (2018) reported that exposing learners of English as a foreign language to the same input (i.e., a TED Talk video) before and after an oral performance (i.e., orally summarize the TED Talk) could successfully facilitate vocabulary learning at the level of meaning recall. It was argued that such improvement could be attributed to the fact that the repeated input had triggered learners' need to make a comparison between their use of target words in the output and how those words were used in the repeated input. While Nguyen and Boers's study is the only one that has suggested that the input-output-cycle could foster vocabulary acquisition, we argue that this technique might help learners perform better in lexical use and fluency in the repeat oral performance as its cognitive mechanisms seem useful for prompting learners' reflection – a fundamental element of L2 oral development (Lynch, 2018). However, this hypothesis has yet to be verified empirically.

2.4. Individual differences as moderating factors

Individual differences have been consistently shown to affect how well L2 learners process and produce language (see Pawlak, 2021; for a review). Among these, working memory capacity and vocabulary knowledge have been demonstrated to play a role in L2 learners' speaking ability, as shown in the following sections.

2.4.1. Working memory

Baddeley (2003, p.189) defines working memory (WM) as a cognitive process that entails "temporary storage and manipulation of information activities that is assumed to be necessary for a wide range of complex cognitive activities." WM has been shown to be useful for regulating learners' attention while processing L2 input (for a review, see Juffs & Harrington, 2011). Fehringer and Fry (2007), using aural WM span test and story retelling test, showed that low WM capacity could result in more hesitations in speaking. In O'Brien et al.'s (2006) study wherein learners were tested the ability to recognize the difference between two series of digits, the results indicated that learners having lower WM tended to speak with less accuracy. Duong and Le (2022) also revealed that L2 learners with higher capacity of storing and processing digits tended to speak faster in a narrative task. This finding, however, should be treated with caution due to the limited scale of the study (i.e., 40 Vietnamese L2 university students). Given that WM seems to play a significant role in L2 input and output processing, there is reason to expect that WM might influence how input exposure and task repetition affect L2 lexical use and fluency.

2.4.2. Vocabulary knowledge

Latest studies have shown the associations between L2 learners' vocabulary knowledge (productive and receptive) and their speaking outcome in terms of fluency and lexical use. For example, Enayat and Derakhshan (2021) demonstrated that L2 university learners' productive vocabulary knowledge, measured by Meara and Fitzpatrick's (2000) Lex30, was a good predictor of fluency. They also found a positive relationship between learners' receptive knowledge and lexical use. Yet, the learners' English proficiency was unclear. Also using Lex 30 test, Clenton et al. (2020) revealed a significant correlation between L2 learners' productive vocabulary knowledge and fluency (i.e., the number of silent pauses) but did not clarify whether the same correlation was found for lexical use aspects. Also, the number of participants was limited (i.e., 30 pre-intermediate Japanese learners of English). Using Nguyen and Nation's (2011) *Vocabulary Size Test*, Duong and Le (2022) revealed that L2 learners with higher receptive vocabulary knowledge tended to use more diverse words when they repeated the same narrative.

2.5. Research rationales and research questions

Some research gaps can be identified based on the literature review. First, while input exposure could lead to L2 lexical mining, it is unknown how input exposure affects other aspects of L2 lexical use and fluency. Further, how input repetition extends the impact of task repetition on lexical use and fluency in the oral repeat task remains unexplored. Second, evidence on the effect of task repetition on L2 lexical use and fluency remains mixed. Finally, findings on the link between working memory/prior vocabulary knowledge and L2 lexical use and fluency remain limited. To address these issues, the following research questions were formulated:

RQ1: To what extent does input exposure affect L2 learners' lexical use and fluency in an immediate and a repeat oral task?
RQ2: Is there an effect of task repetition on L2 learners' lexical use and fluency in the repeat oral task? If so, to what extent?
RQ3: To what extent does input repetition affect L2 learners' lexical use and fluency in the repeat oral task?
RQ4: How do L2 learners' working memory and prior vocabulary knowledge moderate the impact of input, input repetition, and task repetition on their lexical use and fluency?

3. Methodology

3.1. Research design

Following Duong et al. (2021b), this study employed a between-subjects design which involves two independent variables: *input* (the between-subjects variable) at three levels (no-input vs. input vs. input repetition) and *task repetition* (the within-subjects variable). Each group received a different treatment: one group was asked to view L2 captioned videos before performing an oral task (i.e., input group), another group was asked to view the same videos before and after performing the oral task (i.e. input repetition group), and one group performed an oral task without watching the videos (i.e., no-input group/baseline group). After two days, all groups performed exactly the same oral task without being informed of their repeat performances. Additionally, they took five tests: two prior vocabulary knowledge tests and three working memory tests. The six dependent variables included two measures of lexical use (i.e., lexical diversity and lexical sophistication) and four measures of fluency (i.e., frequency of pauses, between clauses and within clauses, frequency of total repairs, articulation rate). The moderating variables were scores of the receptive and productive vocabulary tests, and scores of the working memory tests (i.e., forward-span, backward-span, and operation-span tasks).

3.2. Participants

The study included 90 Vietnamese undergraduates (50 females and 40 males, 19-21 years old) who studied English as a foreign language in three different universities.

The participants took part in the study on a voluntary basis. Their English proficiency ranged from 340 to 750 points (*M* = 501.24, *SD* = 93.16) on the Test of English for International Communication (TOEIC), corresponding to A2-B1 level (ETS, n.d.). The participants were randomly allocated to one of three groups: an input group (*N* = 29), an input repetition group (*N* = 32), and a no-input group (*N* = 29).

3.3. Input

We used short English videos (total time = 6 minutes) with captions (i.e., subtitles in English), which were first introduced in Duong et al. (2021a). The videos feature speeches by native speakers describing Icelandic and Cairns tourist attractions. These videos were selected to avoid the effect of contextual knowledge (Miller et al., 2006) due to participants' unfamiliarity with these destinations. The lexical profile of the videos was checked with Nation's Range software which showed that the 3,000 most frequent word families (without proper nouns) provided 95% coverage. Results of the English-Vietnamese Vocabulary Size Test designed by Nguyen and Nation (2011) revealed that the participants could answer approximately 43 correct questions, indicating that they had a receptive vocabulary knowledge of 8,600 word families on average. As Durbahn Quinteros et al. (2019) found, good comprehension of audiovisual input could be achieved at the 87% coverage, the participants' receptive vocabulary test scores indicated that they probably had sufficient comprehension of the videos.

3.4. Oral task

The present study employed the same picture-prompted narrative task used in Duong et al. (2021b). In this task, learners had to manipulate and reorganize the information processed from the videos to create their own story in our study. They were asked to use seventeen images as prompts to describe their (imaginary) trips to Iceland and Cairns. This task type is used because Skehan (2009) proposes that manipulating and organizing information may elicit more complex language. Prompts were used to avoid excessive individual variation in story lines. Participants were given three minutes of planning time, as this was found to be adequate for learners to prepare what they wanted to say and how to put the pre-verbal messages into appropriate words and structures.

3.5. Prior vocabulary knowledge tests

We tested two aspects of participants' prior vocabulary knowledge. Receptive vocabulary knowledge was assessed using the *English-Vietnamese Vocabulary*

Size Test by Nguyen and Nation (2011), which gauges test takers' capability to recall the words' meanings. The original VST contains 140 sample words drawn from 14 frequency bands (10 words from each band) of 1000 words (1K-14K); however, we decided to use the short version with 70 items as this version was shown reliable and valid by Duong et al. (2021a). In the present study, the items were shown internally consistent (Cronbach's alpha = .87, *N* = 90). To measure learners' productive vocabulary knowledge, we used Laufer and Nation's (1999) Productive Levels Test (PLVT) which aims to evaluate test takers' capability to recall word forms at the 2000, 3000, 5000 and 10,000-frequency levels in a sentence completion task. A high level of internal consistency was found (Cronbach' alpha = .94, *N* = 90).

3.6. Working memory tests

Following Duong et al. (2021b), three non-verbal tests were used to assess WM to facilitate the generalizability. In particular, a forward-digit span task was used to assess learners' capacity to store phonological information and a backward-digit span task to evaluate their capability to store and manipulate phonological input. The forward-digit span task had 8 spans, each of which had 2 sets of 3 to 9 digits. The backward-digit span task consisted of 8 spans, ranging from 2 to 8 digits with 2 sets per span. In both tests, participants listened to digit sequences pre-recorded in Vietnamese. After listening, participants reported the sequences in the exact (forward-span) and reverse order (backward-span) to the examiner. Participants could proceed to the next span only if they could correctly recall at least one sequence of digits. The participants' score (max = 16) equals the total number of correctly recalled sets. The Vietnamese-version Ospan task created in Duong et al. (2021b) was used to assess learners' executive and attention-regulated WM functions. Participants were required to perform mathematical operations as well as memorize the Vietnamese word appearing after each operation. The task consisted of 15 sets, each with 2 to 6 operation-word strings. After each complete set, participants had to write the words down in the presented order on an answer sheet.

3.7. Lexical use and fluency measures

We focused on lexical sophistication and lexical diversity – the aspects that have received increasing attention in recent research (e.g., Duong & Le, 2022; Kim et al., 2018; Wang, 2014). Following Kyle (2020), we conceptualized lexical sophistication as the number of words that are less frequently used by learners in their oral output. To assess lexical sophistication, we adopted band-based as well as count-based approaches. The band-based approach considers a text lexically sophisticated if it contains words that belong to the 3,000 or higher frequency level (K3+ level) (i.e., low-frequency words) (Uchihara & Clenton, 2018). The count-based approach calculates corpus-based average frequency per text to evaluate its lexical sophistication (Crossley et al., 2013).

We conceptualized lexical diversity as "the variety of active vocabulary developed by a speaker" (Malvern & Richards, 2002, p. 87) and used HD-D index to measure lexical diversity. As this index is not influenced by the length of the text (Kyle, 2020), it is considered more reliable in reflecting lexical diversity in spoken discourse than other indices (e.g., MTLD, vocD). This index "represents the sum of probabilities that each type in a text will occur in a 42-word sample at least once" (Kyle, 2020, p. 460).

We adopted Skehan's (2003) framework to analyze learners' fluency. The framework views L2 fluency in terms of three dimensions: speed, breakdown, and repair fluency. Specifically, speed fluency was determined by learners' speed to articulate content words in a speech (Lambert et al., 2020). Breakdown fluency was determined by studying the frequency of silent and filled pauses between and within clauses. Following de Jong and Bosker (2013), a silent pause refers to any period of silence lasting at least 250 ms. Filled pauses were defined as fillers such as *um* and *uh* (Kang et al., 2010). A clause in speech contains at least one clause element, such as a subject, object, or complement, with a finite or nonfinite verb (Foster et al., 2000). The last dimension, repair fluency, reflects the monitoring and repair processes such as self-corrections and repetitions (Tavakoli & Wright, 2020).

3.8. Procedure

We collected data in three one-to-one sessions. In the first session, the input and input repetition group were asked to watch the videos twice on a computer screen (without taking notes) and then perform the immediate oral task. Before watching the videos, they were briefed on the follow-up oral task of the same topic. Participants could not take notes or rewind the videos during the viewing process. They were then given three minutes to prepare for the talk with note-taking allowed. Their notes were collected right after they completed the oral task. Following the immediate oral task, the input group was dismissed, while the input repetition group was instructed to rewatch the videos one more time without note-taking and tell the teacher which location they preferred. Participants in the no-input group completed only the oral task and did not watch any videos. Following the speaking activity, each of the three groups took a 15-minute break before taking the forward and backward-digit span tests (about 15 minutes in total). Two days later, in the second session, all three groups did not watch the videos but were asked to complete the same oral task in three minutes. Before doing the task, they had 15 minutes to note down main ideas. The session ended with all participants doing the Ospan task in about 30 minutes. Two prior vocabulary knowledge tests were completed in the third session with a 15-minute break in between.

3.9. Data analysis

Each speech was audio-recorded and then transcribed manually by the first author then verified by the third author. The proportion of K3+ in the texts were analyzed using Vocab Profile in the Complete Lexical Tutor (available at https://www.lextutor.ca/vp/comp). The transcripts were lemmatized prior to the analysis under the assumption that participants already had knowledge of the verb and noun inflection system. Prior to calculation, verbatim repetitions and obvious pronunciation errors (e.g., "the pear" instead of "the spear") were also fixed. Next, we used the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015) to calculate the logarithm transformed average frequency scores (in reference to the SUBTLEXus Corpus) (Brysbaert & New, 2009) to minimize Zipfian distributions (i.e., disproportional frequency caused by very frequent words) as suggested by Kyle (2020). We used the TAALED (Tool for Automatic Analysis of Lexical Diversity) by Kyle et al. (2020) to calculate the HD-D index with word types as the lexical units. Texts with a lower HD-D value were considered more diverse than those with a higher one (Kyle, 2020).

We used Praat 6.1.42 (Boersma & Weenink, 2012) and Syllable Nuclei Praat script (de Jong & Wemp, 2009) to calculate the phonation time, the quantity of syllables, and the number of silent pauses in utterances contributing to the task content. This means utterances at the beginning and the end of performances (e.g., "Hello, my name is …" or "That's all" or "Thank you for your listening") were excluded. We observed that there was a chance that some syllables were missed by the software; thus, we double-checked by listening to the audios while examining the Praat spectoram to note down non-recognized syllables on Praat grid.

The total number of filled pauses was then manually coded. Silent and filled pauses were coded as follows: 1) silent pauses within clauses, 2) silent pauses between clauses, 3) filled pauses within clauses, and 4) filled pauses between clauses. The number of repairs was also counted manually. The first and third authors performed independent coding. 20% of the transcripts were coded first. High interrater reliability was established on the five variables: silent pauses within clauses (r = .998), silent pauses between clauses (r = .992), filled pauses within clauses (r = .987), filled pauses between clauses (r = .987), and repairs (r = .994). The remainder of the transcripts was then exclusively coded by the first author. After that, articulation rate equals the total number of syllables divided by the phonation time. The frequency of silent and filled pauses per minute was determined by dividing the number of silent and filled pauses by the phonation time multiplied by 60. The frequency of repairs was also calculated in a similar way, ((total number of repairs/phonation time) x 60). The frequency calculation was used to control for the length of performances, which varies across the participants.

3.10. Statistical analysis

The results were analyzed with SPSS 27 and R software (version 4.0.1, R Core Team, 2020). It was found that the HD-D scores, K3+ proportion, and log-transformed average frequency scores were normally distributed. Yet, the distribution of pauses frequency and repairs frequency was positively skewed. Following Lambert et al. (2020), we combined filled and unfilled pauses between and within clauses into composite measures of total pauses between clauses and within clauses. We found that it was possible to normalize the distributions of pauses frequency between clauses, within clauses, and repair frequency with Log10 transformation.

Repeated measures correlation tests using rmcorr function (version 0.4.6) in R software (Bakdash & Marusich, 2017) revealed a significant correlation, albeit weak,[1] between the two measures of lexical sophistication (e.g., average frequency and proportion of K3+ lemmas) and moderate correlations among the four measures of fluency in the immediate task and the repeat task (see Appendix for the correlation matrix). Therefore, we decided to compute a repeated measures ANCOVA for lexical diversity, a repeated measures MANCOVA for lexical sophistication measures and a repeated measures MANCOVA for fluency measures with Bonferroni post-hoc tests.

For each model, all participants completed an immediate oral task and a repeat oral task; thus, *task repetition* (immediate vs. repeat) was a within-subjects factor. *Input exposure* (input vs. input repetition vs. no input) was a between-subjects factor. Covariates were scores of the working memory tests and the prior vocabulary knowledge tests. Since all covariates displayed Pearson correlations below .70, the threshold for multicollinearity effect to occur (e.g., Crossley et al., 2011), they were all included in the regression models. Other assumptions including outlier bias, homogeneity of variance-covariance were also checked for each model. Examination of histograms and boxplots revealed that there were no outliers in all models. Levene's tests for equality of error variance were not significant ($p > .05$), indicating that the dependent variables' error variances were equal across the groups. An alpha level of .05 in Wilk's lambda test was taken as the level of statistical significance.

[1] Cohen's (1988) interpretation is used for these r values: .1-.3: small effect, .3-.5: medium or intermediate effect, .5 and higher: large or strong effect.

4. Results

4.1. Participants' individual differences

Table 1 reveals that participants in all three groups performed similarly in the receptive vocabulary test and productive levels test. An ANOVA showed that the differences were not significant: receptive test ($F(2, 87) = .195$, $p = .823$) and productive test ($F(2, 87) = 2.784$, $p = .067$).

Table 1 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)	38.04-45.32	28.18 (15.33)	22.60-33.76
Input (*N* = 29)	43.36 (7.93)	40.61-46.11	31.57 (12.01)	27.41-35.73
No input (*N* = 29)	43.95 (7.79)	41.11-46.79	35.59 (8.55)	32.48-38.70

Participants were shown to have good and comparable scores on the working memory tests (see Table 2). No significant differences were found amongst the three groups: forward-span ($F(2, 89) = .822$, $p = .443$), backward-span ($F(2, 89) = .249$, $p = .780$), and Ospan task ($F(2, 89) = 1.43$, $p = .243$).

Table 2 Descriptive statistics for the three working memory tests

	Forward-span (max = 16) Mean (*SD*) 95% CI	Backward-span (max = 16) Mean (*SD*) 95% CI	OSPAN (max = 60) Mean (*SD*) 95% CI
Input repetition (*N* = 32)	14.43 (2.09)	14.06 (2.50)	50.86 (8.43)
	13.67-15.20	13.15-14.97	47.79-53.93
Input (*N* = 29)	13.50 (3.45)	13.59 (2.78)	51.18 (11.13)
	12.30-14.70	12.63-14.55	47.95-55.67
No input (*N* = 29)	14.10 (2.90)	13.96 (3.05)	46.58 (14.76)
	13.04-15.16	12.85-15.07	41.21-51.95

4.2. The impact of input, input repetition, and task repetition on lexical use and fluency of L2 learners

Table 3 displays descriptive statistics for all lexical use and fluency measures. The repeated measures ANCOVAs for lexical diversity and repeated measures MANCOVAs for lexical sophistication did not show a main effect of input and task repetition on all measures of lexical use and fluency in the immediate as well as the repeat task. The repeated measures ANCOVA for lexical diversity only revealed an interaction effect between input and task repetition with a large effect size[2] (F(2, 82) = 5.377, p = .006, η_p^2 = .116). Specifically, Figure 1 shows that while the no-input group's word choice seemed wider (i.e., lower HD-D index) in the repeat than the immediate task, word choice in the input and input repetition groups seemed narrower. This finding suggests that input exposure, repeated or not, may constrain learners' ability to use more diverse words as they repeat the oral task.

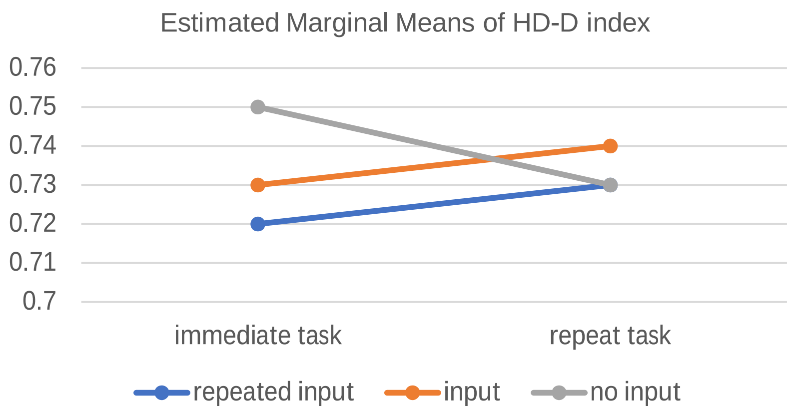

Figure 1 Interaction effect of input exposure and task repetition on lexical diversity

Table 3 Descriptive statistics of oral performance measures per groups

	Groups	Immediate task		Repeat task	
		Mean (*SD*)	95% CI	Mean (*SD*)	95% CI
HD-D	NI	0.75 (0.05)	0.73-0.77	0.73 (0.03)	0.72-0.74
	I	0.73 (0.02)	0.72-0.74	0.74 (0.03)	0.73-0.75
	IR	0.72 (0.05)	0.70-0.74	0.73 (0.04)	0.72-0.74
Log-transformed average frequency	NI	3.85 (0.17)	3.79-3.91	3.95 (0.52)	3.76-4.14
	I	3.90 (0.15)	3.85-3.95	3.87 (0.13)	3.82-3.92
	IR	3.88 (0.16)	3.82-3.94	3.90 (0.15)	3.85-3.95
K3+ proportion	NI	11.20 (4,08)	9.72-12.68	11.58 (3.56)	10.28-12.86
	I	12.50 (4.01)	11.11-13.89	12.28 (3.74)	10.92-13.64
	IR	13.50 (2.90)	12.45-14.56	12.84 (3.12)	11.70-13.98
Articulation rate per second	NI	3.49 (0.42)	3.34-3.64	3.63 (0.49)	3.45-3.81
	I	3.37 (0.53)	3.19-3.55	3.21 (0.54)	3.01-3.41
	IR	3.16 (0.42)	3.01-3.31	3.09 (.048)	2.92-3.26
Within-clause pauses per minute	NI	20.97 (4.35)	17.57-24.37	20.84 (5.15)	16.42-25.26
	I	22.71 (3.39)	20.15-25.27	21.70 (3.91)	18.46-24.94
	IR	21.17 (3.84)	17.95-24.39	22.74 (3.10)	19.43-26.01
Between-clause pauses per minute	NI	13.52 (4.18)	12.00-15.04	15.38 (4.67)	13.68-17.08
	I	14.18 (4.09)	12.76-15.60	15.26 (5.04)	13.43-17.09
	IR	13.13 (2.72)	12.14-14.12	14.51 (4.71)	12.80-16.22
Repairs per minute	NI	4.93 (3.46)	3.67-6.19	3.78 (2.89)	2.73-4.83
	I	2.52 (1.63)	1.96-3.08	4.07 (3.94)	2.64-5.50
	IR	2.94 (2.87)	1.90-3.98	3.36 (3.17)	2.21-4.51

Note. NI: No-input, I: Input, IR: Input repetition

[2] *Note*. An effect size (η_p^2) of .01, .06 and .14 or above was considered small, medium and large, respectively.

The repeated measures MANCOVA revealed a negative effect of input (F(8, 160) = 3.488, p = .001) on overall fluency. Repeated measures ANCOVAs for each fluency construct revealed that the effect of input was found only on articulation rate and the effect was only significant in the repeat task (p < .001) (see Table 5). Post-hoc pairwise comparisons displayed a significant difference between the no-input and input group (p < .001) as well as between the no-input and the input repetition group (p = .005). Together with the descriptives, these findings seem to indicate that input exposure had a negative effect on learners' articulation rate especially in the repeat task. The post hoc comparisons did not show a significant difference between the input and input repetition groups (p = .793), indicating that repeated input did not play a significant role. The impact of task repetition on fluency measures was not found.

4.3. The influence of working memory and vocabulary knowledge

To examine the mediating effect of individual differences, particularly prior vocabulary knowledge as well as working memory, we used separate one-way ANCOVAs, with adjusted p-values. For lexical use, no link was found between learners' prior vocabulary knowledge/working memory and HD-D index as well as average frequency scores. However, ospan task scores seemed to be a good predictor for K3+ proportion in the repeat task (p = .002, β = .273) (see Table 4). This finding indicates that learners with greater storage and executive and attention-regulated WM capacity tended to use more low-frequency words in the repeat task, but the predictive power is limited as evidenced by the small β value.

Table 4 One-way ANCOVAs for K3+ proportion in the immediate and repeat task

		K3+ proportion				
		df	*F*	*p*	$\eta 2$	Power
Input	Immediate	2	1.038	.359	.025	.226
	Repeat	2	.055	.946	.001	.058
Receptive	Immediate	1	.425	.516	.005	.099
	Repeat	1	.472	.494	.006	.104
Productive	Immediate	1	6.156	.015	.070	.689
	Repeat	1	1.455	.231	.017	.222
Forward-span	Immediate	1	5.825	.018	.066	.665
	Repeat	1	.436	.511	.005	.100
Backward-span	Immediate	1	.732	.395	.009	.135
	Repeat	1	1.258	.265	.015	.198
Ospan	Immediate	1	1.986	.162	.024	.286
	Repeat	1	10.530	.002	.114	.894
Error	Immediate	82				
	Repeat	82				

Note. *p*-value was set at 0.0125

For fluency measures, the analysis showed a significant relationship between learners' receptive vocabulary size and the frequency of pauses within the clauses (*p* = .003, *β* = -.009) in the repeat task (see Table 5). This finding indicates that participants with larger vocabulary size tended to pause less than those with smaller lexicon. In the repeat task, the analyses showed correlations between the backward-span test scores and learners' articulation rate (*p* = .001, *β* = .071) and frequency of pauses between clauses (*p* = .002, *β* = .018). This indicates that learners having greater capacity of storing and processing information articulated faster and paused more often at the end of the clauses. No link was found between prior vocabulary knowledge, working memory and repairs fluency.

Table 5 One-way ANCOVAs of fluency measures in the immediate and repeat task

		df	Articulation rate				Pauses within clauses				Pauses between clauses				Repairs			
			F	*p*	*η*2	Power	*F*	*p*	*η*2	Power	*F*	*p*	*η*2	Power	*F*	*p*	*η*2	Power
Input	Immediate	2	4.844	.010	.106	.787	.613	.544	.015	.149	.012	.988	.000	.052	3.734	.028	.083	.669
	Repeat	2	9.619	.000	.190	.978	1.330	.270	.031	.280	.213	.808	.005	.082	.362	.698	.009	.106
Receptive	Immediate	1	2.366	.128	.028	.330	6.957	.010	.078	.741	.222	.639	.003	.075	2.305	.133	.027	.323
	Repeat	1	3.219	.076	.038	.426	9.073	.003	.100	.845	1.587	.211	.019	.238	1.544	.084	.036	.409
Productive	Immediate	1	.196	.660	.002	.072	4.257	.042	.049	.532	2.894	.093	.034	.390	1.172	.282	.014	.188
	Repeat	1	.011	.918	.000	.051	3.735	.057	.044	.480	3.415	.068	.040	.447	.342	.560	.004	.089
Forward-span	Immediate	1	1.304	.257	.016	.204	1.035	.312	.012	.171	1.808	.182	.022	.265	.667	.416	.008	.127
	Repeat	1	.979	.325	.012	.165	.090	.765	.001	.060	1.286	.260	.015	.202	.100	.753	.100	.061
Backward-span	Immediate	1	3.326	.072	.039	.437	.122	.727	.001	.064	6.770	.115	.076	.730	.991	.322	.012	.166
	Repeat	1	12.137	.001	.129	.931	1.369	.245	.016	.212	9.851	.002	.107	.873	1.034	.312	.012	.171
Ospan	Immediate	1	.403	.528	.005	.096	.171	.681	.002	.069	.863	.356	.010	.151	.019	.891	.000	.052
	Repeat	1	.076	.784	.001	.059	1.754	.189	.021	.258	.009	.963	.001	.056	.629	.430	.008	.123
Error	Immediate	82																
	Repeat	82																

Note. *p*-value was set at 0.00625

5. Discussion

The current study extends previous research by showing that L2 learners' lexical use and fluency in speaking can be affected by input exposure, input repetition, and task repetition. The findings will be discussed in more details in section 5.1, 5.2, and 5.3, respectively.

5.1. The impact of input exposure on L2 lexical use and fluency in an immediate and a repeat oral task (RQ 1)

For the immediate task, we did not find an impact of meaningful input exposure on lexical use and fluency for the immediate task. This indicates that Bygate's (1996) and Willis's (1996) hypotheses, predicting that exposing learners to input relevant to output tasks can lead to improved lexical use and fluency, cannot be confirmed. The absence of the effect of input on the immediate task might be explained by the learners' low proficiency level. At this level of proficiency (A2-B1), learners' lexical resources may still be limited to high-frequency words. It is also possible that the small proportion of low-frequency words in the videos (i.e., 3.83%) may have limited learners' willingness to take a risk of using more low-frequency words in the oral task. Another possibility is learners' processing of words from input, as shown in Duong et al. (2021b), may have spared them with less processing time to recall low-frequency words and widen their word choice. Therefore, they might have tried to buy time by slowing down the speech rate to alleviate the cognitive burden. For the repeat task, the input group tended to produce words with less variety and speak more slowly than those in the no-input group. We conducted a follow-up analysis which revealed a considerable proportion of verbatim repetition[3] (approximately 46%) from the first to the repeat task performance for the no-input group compared to 28.61% for the input repetition group and 27.97% for the input group. These findings indicate that verbatim repetition might have aided participants in the no-input group to speed up their speaking rate, thus providing them with more time to widen their word choice.

5.2. The impact of task repetition and input repetition in L2 lexical use and fluency in a repeat task (RQ2 and RQ3)

We found that the effect of task repetition on L2 learners' lexical sophistication at the second repetition was not significant. This finding is similar to Duong and Le (2022). As Gass et al. (1999) and Kim et al. (2018) only found the significant effect of task repetition after three times of repetition, this might indicate that learners should have more time one time of repetition to improve their lexical use. Unlike Fukuta (2016), our study did not reveal an impact of task repetition on lexical diversity at the second repetition, which might be due to discrepancies in the measures of lexical diversity (i.e., HD-D index vs. Guidraud) and time between tasks (i.e., one week vs. two days). Contrary to a widely held belief that task repetition could promote fluency (Bygate, 2018), our study did not show an impact of task repetition on learners' fluency, which is consistent with Duong and Le (2022). As the participants of both studies are Vietnamese EFL learners and did not follow a task-based language program, replication studies with different learners' backgrounds are needed.

Concerning the effect of input repetition, while exposing learners to the same input before and after performing speaking task was useful for them to learn the meaning of new words from input (Nguyen & Boers, 2018) and use newly

[3] As a follow-up analysis, verbatim repetition was generally calculated by counting the same words across the repetitions of oral task, using Text Lex Compare function at lextutor.ca. It should be noted that previous studies (e.g., Thai & Boers, 2016) adopted a more fine-grained approach to calculate verbatim repetition.

learned words in follow-up oral task (Duong et al., 2021b), this did not seem to benefit learners' lexical use and fluency in the repeat speaking task. It is possible that exposing learners to the same input twice might be insufficient to familiarize them with the topic or might not provide them with adequate time to enrich their lexical use or speed up their speaking rate. Also, it cannot be excluded that other forms of input repetition (i.e., repeating input before the output task) might result in different findings. This issue warrants further research.

5.3. The influence of prior vocabulary knowledge and working memory (RQ 4)

The findings suggest that we can somehow predict learners' fluency and lexical sophistication if we know their receptive vocabulary knowledge and WM, but the predictive power is limited. Specifically, learners with larger receptive vocabulary size tended to pause less within the clauses, which is not in line with de Jong (2016) and Clenton et al. (2020) who found a link between learners' pausing and productive vocabulary knowledge. Further, unlike Uchihara and Saito (2019) and Clenton et al. (2020), we did not find a significant correlation between productive vocabulary and lexical sophistication. We assume that the difference in the test format between our study and others might have played a role in the difference in findings. While we used the Productive Vocabulary Levels Test (PVLT) which tests the form recall knowledge, de Jong (2016) used a Dutch-version sentence completion test and Clenton et al. (2020) administered Lex30 which measures word associations. Also, as suggested by Clenton et al. (2020), it is possible that the scores achieved from the PVLT might be more aligned to the ability to use words by participants at the advanced rather than those at low-intermediate proficiency levels such as the learners in our study. Our finding is also not in line with Duong and Le (2022) who reported a positive link between Vietnamese learners' receptive vocabulary knowledge and lexical diversity. As Duong and Le (2022) employed a more spontaneous oral task without prompts, their participants might have had more room to widen their vocabulary choice while performing the oral task than our participants.

Concerning WM, the results show that learners with greater Ospan task scores used more low-frequency words, while those with greater backward-span task scores spoke slightly faster and paused more often at the end of clauses in the repeat task. These findings indicate that executive and attention-regulated WM can predict not only vocabulary acquisition (e.g., as in French, 2014; Montero Perez, 2020) but also lexical use and fluency. Our study demonstrates that forward digit span scores were not predictors of L2 learners' lexical use and fluency, which seems to support findings reported in Juffs and Harrington's (2011) meta-analysis that phonological WM is often less important for learners beyond low-proficiency level.

6. Conclusion

The present study has investigated the effects of input, input repetition and task repetition on L2 learners' lexical use and fluency. The study also explored to what extent learners' lexical use and fluency in the immediate and the repeat oral task is predicted by their prior vocabulary knowledge and working memory. The findings indicate that input, input repetition, and task repetition do not improve learners' lexical use and fluency in either the immediate or repeat task. Working memory and prior vocabulary knowledge can predict L2 learners' lexical sophistication and fluency in the repeat task, but the association is not strong.

The present study has some limitations. First, as the study focused on university students, it is difficult to generalize the findings to participants of other profiles. Second, because the repeat task was administered only two days after the initial task, readers should be cautious in generalizing the findings to scenarios where tasks are repeated more than once. Third, as our study only focused on picture-prompted oral narrative tasks and input repetition after output, future studies that use other types of output tasks and other forms of input repetition (e.g., repeating input before output) are needed to generalize the findings.

From a pedagogical perspective, our findings suggest that the combination of task repetition and input exposure does not seem a judicious option if the sole aim of the activity is to foster immediate improvements in oral fluency and lexical use. Remarkably, the predictive role of receptive vocabulary knowledge and working memory for lexical use and fluency indicates that individual differences might need more attention in L2 training programs. For example, learners at low proficiency levels might need extra support from teachers to extend their lexical repertoires before being trained for better fluency and lexical use. Also, learners who have difficulties in remembering information (e.g., the elderly learners, busy adult learners) might need to have additional time for task preparation to compensate for their limited capacity of storing and processing information.

References

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

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology, 8*(456). https://doi.org/10.3389/fpsyg.2017.00456

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990. https://doi.org/10.3758/s13428-013-0403-5

Boersma, P., & Weenink, D. (2012). PRAAT. Retrieved May 12, 2012 from http://www.praat.org

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

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). MacMillan Heinemann.

Bygate, M. (2018). Introduction. In Bygate, M. (Ed.), *Learning language through task repetition* (pp. 1-25). John Benjamins.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in second language* (pp. 37-74). John Benjamins.

Clenton, J., de Jong, N. H., Clingwall, D., & Fraser, S. (2020). Vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2). In Clenton, J. & Booth, P. (Eds.), *Vocabulary and the four skills: Pedagogy, practice, and implications for teaching vocabulary* (pp. 146-165). Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Publishers.

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

Crossley, S. A., Cobb, T., McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System, 411*(4), 965-981. https://doi.org/10.1016/j.system.2013.08.002

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *IRAL-International Review of*

Applied Linguistics in Language Teaching, 54(2), 113-132. https://doi.org/10.1515/iral-2016-9993

De Jong, N. H., & Wemp, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385-390. https://doi.org/10.3758/BRM.41.2.385

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS)* (pp. 17-20). Royal Institute of Technology.

Durbahn Quinteros, M., Rodgers, M., & Peters, E. (2019). The relationship between vocabulary and viewing comprehension. *System*, *88*, 1-38. https://doi.org/10.1016/j.system.2019.102166

Duong, P.-T., Montero Perez, M., Desmet, P., & Peters, E. (2021a). Learning vocabulary in spoken input- and output-based tasks. *TASK 1*(1), 100-126. https://doi.org/10.1075/task.00005.duo

Duong, P.-T., Montero Perez, M., Nguyen, L.-Q., Desmet, P., & Peters, E. (2021b). Incidental lexical mining in task repetition: The role of input, input repetition and individual differences. *System, 103*, 1026-1050. https://doi.org/10.1016/j.system.2021.102650

Duong, P.-T., & Le, H. H. V. (2022). How working memory and prior vocabulary knowledge influence the impact of task repetition on L2 oral performance: Insights into Vietnamese EFL learners. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, *26*(3). https://doi.org/10.55593/ej.26103a12

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Enayat, M. J., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*. https://doi.org/10.1016/j.system.2021.102521

Fehringer, C., & Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica: Acta Societatis Linguisticae Europapeae, 41*(1-2), 37-72. https://doi.org/10.1515/flin.41.1-2.37

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

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354-375. https://doi.org/10.1093/applin/21.3.354

Gass, S., Macket, A., Alvarez-Torres, M., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49(*4), 549-581. https://doi.org/10.1111/0023-8333.00102

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

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

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*, 554-566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Khatib, M., & Farahanynia, M. (2020). Planning conditions (strategic planning, task repetition, and joint planning), cognitive task complexity, and task type: Effects on L2 oral performance. *System, 93*. https://doi.org/10.1016/j.system.2020.102297

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). John Benjamins.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786. https://doi.org/10.1002/tesq.194

Kyle, K. (2020). Measuring lexical richness. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 454-476). Routledge.

Kyle, K., Crossley, A. S., & Jarvis, S. (2020). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly, 18*(2), 154-170. https://doi.org/10.1080/15434303.2020.1844205

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322. https://doi.org/10.1093/applin/16.3.307

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

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39*(1), 167-196. https://doi.org/10.1017/S0272263116000085

Lambert, C., Aubrey, S., Leeming, P. (2020). Task preparation and second language speech production. *TESOL Quarterly*. https://doi.org/10.1002/tesq.598

Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT Press.

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). John Benjamins.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85-104. https://doi.org/10.1191/0265532202lt221oa

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System, 28*, 19-30. https://doi.org/10.1016/S0346-251X(99)00058-5

Miller, L. M. S., Cohen, J. A., & Wingfield, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory & Cognition, 34*, 1355-1367. https://doi.org/10.3758/bf03193277

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, 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

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

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 Psycholinguistics, 27*, 377-402. https://doi.org/10.1017/S0142716406060322

Pawlak, M. (2021). *Investigating individual learner differences in second language learning*. Springer.

R Core Team (2020). R: A language and environment for statistical computing (Version 4.0.1) [Computer software]. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

Skehan, P. (2003). Task based instruction. *Language Teaching*, *36*, 1-14. https://doi.org/10.1017/S026144480200188X

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510-532. https://doi.org/10.1093/applin/amp047

Tavakoli, P., & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge University Press.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly, 50*(2), 369-393. https://doi.org/10.1002/tesq.232

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

Wang, Z. (2014). Online time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing Perspectives on Task Performance* (pp. 27-61). John Benjamins.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225-239). Routledge.

Willis, J. (1996). *A framework for task-based learning*. Addison-Wesley.

APPENDIX

The correlation matrix of repeated measures correlation coefficiencies

	HD-D	Average frequency	K3+ proportion	Articulation rate	Within-clause pauses	Between-clause pauses	Repairs
HD-D		-.05 (*p* = .62)	.03 (*p* =.70)	.02 (*p* = .85)	-.15 (*p* = .14)	.04 (*p* = .63)	.08 (*p* = .41)
Average frequency	-.05 (*p* = .62)		.27** (*p* = .009)	.12 (*p* = .25)	.09 (*p* = .38)	.25* (*p* = .016)	-.17 (*p* = .09)
K3+ proportion	.03 (*p* - .70)	.27** (*p* = .009)		-.08 (*p* = .41)	.20 (*p* = .056)	-.11 (*p* = .27)	.19 (*p* = .06)
Articulate rate	.02 (*p* = .85)	.12 (*p* = .25)	-.08 (*p* = .41)		.43** (*p* < .001)	.53** (*p* < .001)	.09 (*p* = .37)
Within-clause pauses	-.15 (*p* = .14)	-.09 (*p* = .38)	.20 (*p* = .056)	.43** (*p* < .001)		.40** (*p* < .001)	.10 (*p* = .32)
Between-clause pauses	.04 (*p* = .63)	.25* (*p* = .016)	-.11 (*p* = .27)	.53** (*p* < .001)	.40** (*p* < .001)		-.08 (*p* = .41)
Repairs	.08 (*p* = .41)	-.17 (*p* = .09)	.19 (*p* = .06)	.09 (*p* = .37)	.10 (*p* = .32)	-.08 (*p* = .41)	

Note. ** Correlation is significant at the 0.01 level; * Correlation is significant at the 0.05 level