Learning vocabulary in spoken input- and output-based tasks

Phuong Thao Duong,[1,3] Maribel Montero Perez,[2]
Piet Desmet,[1,3] and Elke Peters[3]
[1] KU Leuven, imec research group itec, Belgium | [2] Ghent University,
Belgium | [3] KU Leuven, Belgium

This experimental study explores the differential effects of spoken input-based and output-based tasks on vocabulary knowledge. The study also investigates whether such tasks result in more learning gains than exposure to input-only (no subsequent task). The study employed a pretest-posttest design with two groups: an experimental group ($n = 32$) who completed both input- and output-based tasks in a counterbalanced way and a comparison group ($n = 12$) who were only exposed to L2 input. Vocabulary gains were measured at three levels of sensitivity: oral spontaneous use, oral form recall and meaning recall. The findings showed that participants who were only exposed to L2 input learned significantly fewer words than participants who completed the input-based and output-based tasks. No difference in learning gains was found between the input-based and output-based tasks.

Keywords: task-based language teaching, input-based tasks, output-based tasks, vocabulary acquisition, audiovisual input

Introduction

Input-based tasks (i.e., tasks in which learners process input via listening or reading and do not have to produce language in the L2) have received growing attention for their potential in improving language learning since Ellis, Tanaka, and Yamazaki (1994). Input-based tasks have been typically operationalized as *listen-and-do* tasks where learners listen to verbal input and demonstrate their comprehension through non-verbal tasks (e.g., listen and choose the correct pictures). Input-based tasks have been investigated on their own (e.g., Shintani, 2012; Erlam & Ellis, 2018) or compared with the Present-Practice-Produce approach (PPP) (e.g., Shintani, 2011, 2013, 2015), but less so with other task-based activities like

output-based tasks (i.e., tasks in which learners are engaged in producing meaningful spoken or written output in the L2). An exception is Duong et al. (in press), who compared the effects of input- and output-based tasks on L2 vocabulary learning as measured by means of written vocabulary tests. It was found that input-based tasks were more suited to learning the meaning of words, while output-based tasks were more beneficial for giving the L2 form. An open question is whether the results would be the same if the tasks and tests were performed in the spoken mode (i.e., speaking tasks and spoken vocabulary tests which require learners to say the responses). It is important to examine this issue because speaking and writing are assumed to pose different demands on cognitive involvement, which might lead to differences in language acquisition (Halliday, 1989; Manchón & Williams, 2016). Additionally, while several vocabulary learning studies have explored the effects of input with follow-up word-focused activities in comparison with input-only without follow-up activities on vocabulary learning (e.g., Hill & Laufer, 2003; Mason & Krashen, 2004; Min, 2008; Peters, Hulstijn et al., 2009), there are few task-based studies that compare the effects of input followed up by tasks with input-only without any follow-up tasks. To fill these gaps, our study investigates the differential effects of input- and output-based tasks on L2 vocabulary learning in the spoken mode. In addition, we compared the effects of input-only without follow-up tasks with input- and output-based tasks to investigate whether the follow-up task contributed to the effects of meaningful input on L2 vocabulary learning.

Background

Within task-based language teaching (TBLT), a 'task' is commonly conceptualized as a meaning-focused activity in which meaning is primary, learners' linguistic resources are not restricted, there is some kind of information gap between interlocutors, and linguistic outcome is not the main task outcome (Ellis, 2003). However, a task can also be form-focused, i.e., designed to provide learners with opportunities to use a specific linguistic form (e.g., vocabulary, grammar) in the task. While numerous TBLT studies have focused on grammar (for a review, see Ellis, 2003), fewer studies have focused on vocabulary learning (Duong et al., in press; Erlam & Ellis, 2018; Shintani, 2011, 2013, 2015). Additionally, output-based tasks seem to have received more attention from TBLT researchers, viz. 85 output-based studies (for a review, see Plonsky & Kim, 2016) than input-based tasks. In the next sections, we will review the literature related to the role of input- and output-based tasks in L2 vocabulary learning. Given that our focus is on the spo-

ken mode, we will discuss mainly studies examining oral (input- and output-based) tasks.

Output-based tasks

Swain's (1998, 2005) Output Hypothesis argues that exposure to input only is insufficient for learners to achieve nativelike L2 competence because of the lack of opportunities for language production and focus on form. In light of that theoretical argument, a wealth of task-based studies has examined the role of output-based tasks in L2 learning and revealed that grammar learning can be improved by output-based tasks (for a review, see Plonsky & Kim, 2016). The few studies that have focused on vocabulary learning explored the effects of output-based tasks after learners had been exposed to L2 input. For instance, Kim (2008) measured intermediate ESL learners' vocabulary gains after they had listened to an L2 text containing target single words and orally reconstructed the text in the L2. Nassaji and Tian (2010) conducted an experimental study with low-intermediate ESL learners who performed oral output tasks in L2 after listening to an L2 text and doing a form-meaning matching exercise. Both studies found that output-based tasks could improve vocabulary knowledge at the level of word meaning.

Recently, Nguyen & Boers (2018) explored the effect of an oral summary task on L2 vocabulary uptake after watching a TED Talk. Sixteen upper-intermediate EFL learners viewed an L2 TED Talk twice, but the experimental group had to orally summarize in the L2 between the first and the second viewing of the talk while the control group did not. The results showed that the experimental group outperformed the control group in the meaning recall test (i.e., L2-L1 translation test). However, given the limited sample size, care should be taken when generalizing these findings. In short, it has been demonstrated that output-based tasks may foster vocabulary learning, but previous research has mainly focused on L2 learners' acquisition of word meaning. Thus, it remains unclear whether output-based tasks have a direct effect on other aspects of vocabulary knowledge such as form and use.

Input-based tasks

Research suggests that input-based tasks may improve vocabulary learning as well (e.g., Erlam & Ellis, 2018; Shintani, 2012). Conceptualized as a type of form-focused tasks, input-based tasks are typically designed to engage learners in input comprehension, and attract their attention to specific linguistic features in a meaningful context (Ellis, 2003). Input-based tasks have frequently been operationalized as *listen-and-do* tasks (i.e., tasks where learners have to listen to verbal

input and demonstrate their comprehension of target language non-verbally). Shintani (2012) explored the effects of listen-and-do tasks on L2 vocabulary acquisition by beginner-level Japanese EFL learners (aged 6–8). The results showed that input-based tasks led to significant gains of target single words at the level of form-meaning connection (i.e., recognize the forms of words and attach a meaning to them) and at the level of form recall (i.e. produce the spoken form of a word). Erlam and Ellis (2018) also explored the effects of listen-and-do tasks with beginner-level learners (aged 13) of French. Their findings mirrored Shintani's (2012), viz. gains of target French single words were found form-meaning connection and form recall levels in the spoken mode. In short, previous research has provided positive evidence for the use of input-based tasks in L2 vocabulary teaching, but the number of studies is still limited and has focused mainly on beginning learners. In addition, Révész (2017) pointed out that more research is needed into other types of input-based tasks beside listen-and-do tasks to determine whether and to what extent findings found in one type can be transferred to other types of input-based tasks.

The effects of input-based tasks have also been compared with present-practice-produce (PPP) activites and output-based tasks (e.g., Duong et al., in press; Shintani, 2011, 2013). Shintani (2011, 2013) compared input-based tasks with PPP activities. PPP activities were operationalized through a set of production-based activities which require learners to produce utterances containing the target linguistic features. Shintani found that the two types of instruction had similar effects on the learning of nouns at the level of spoken form and meaning recall (Shintani, 2011), but input-based tasks were better than PPP for learning the form of adjectives (Shintani, 2013). While Shintani's findings are valuable in revealing how two types of activities influence L2 vocabulary learning, they do not clarify whether similar effects can be found between activities within the same TBLT approach, that is, input-based tasks vs. output-based tasks. Duong et al. (in press) aimed to answer that question by comparing the effects of written input- and output-based tasks on L2 vocabulary learning. It was shown that both task types led to vocabulary learning, but input-based tasks were better than output-based tasks in developing written meaning recall while output-based tasks were more useful in developing written form recall. The findings were explained in light of the Transfer Appropriate Processing (TAP) theory (Morris, Bransford, & Franks, 1977), which claims that learners' test achievement will be at its best when the retrieval processes in the tests match the learning processes in the tasks. However, since Duong et al. used writing tasks and written tests, it is unclear whether the findings also hold for vocabulary knowledge in the spoken mode. According to Schoonen et al. (2009), compared to speaking, the act of writing naturally provides learners with more opportunities to encode form and meaning which can

promote focus on form – a process that can facilitate acquisition as claimed by Swain (1998).These opportunities are afforded by: (1) the slower pace of writing which engages learners in cognitive processes such as planning, noticing, reflecting that encourage them to access and retrieve explicit and/or implicit knowledge in sources of input (e.g. reference materials); and (2) the permanent record left by writing which allows learners to compare retrieved knowledge to their output – the processes that can only be registered fleetingly in speaking (Manchón & Williams, 2016).

L2 vocabulary learning from L2 input and input followed up by activities

Numerous studies have investigated the effects of L2 input-only on incidental vocabulary learning (i.e., learning new words as a by-product of reading, listening, or TV viewing without the intention to learn to a specific word). It has been shown that different types of L2 input, like reading (e.g., Pellicer-Sánchez & Schmitt, 2010), reading-while-listening (Webb et al., 2013), listening (e.g., Vidal, 2011; Jin & Webb, 2020), and TV viewing (e.g., Peters & Webb, 2018; Rodgers & Webb, 2020) are beneficial for vocabulary learning and that learners can learn multiple aspects of word knowledge, like the spelling, meaning, or form (e.g., Pellicer-Sánchez & Schmitt, 2010; Peters & Webb, 2018). However, incidental learning gains tend to be quite modest. Research has shown that input followed by supplementary activities could enhance the effects of L2 input-only, for instance, reading followed by vocabulary activities (e.g., sentence writing with target words, fill-in-the-blank, L1 translation) led to greater vocabulary learning than reading-only (Hill & Laufer, 2003; Min, 2008; Peters et al., 2009) or listening followed by comprehension questions were better than listening-only (Mason & Krashen, 2004). Within TBLT research, Nguyen and Boers (2018) also found greater learning gains for the group that had to do a follow-up task (= L2 oral summary) after watching a TED Talk as compared to the group who only watched the TED Talk without doing the oral summary. However, since Nguyen and Boers used output tasks, it remains to be seen whether similar findings can be found with input-based tasks where learners do not have to produce output in the L2. More research is thus warranted to investigate to what extent input-based tasks can enhance the effects of input-only, as has been suggested by Révész (2017).

Research rationale and research questions

Despite the growing evidence for the potential of input-based tasks for vocabulary learning, empirical studies on input-based tasks are scarce and tend to be restricted to listen-and-do tasks. Further, the effects of input-based tasks have

been mainly compared with PPP activities, with the exception of Duong et al.'s study which compared the effects of input- and output-based tasks on vocabulary learning, but this study focused on writing tasks and written tests. Also, given that there has been considerable evidence regarding how writing and speaking have a differential impact on learners' cognitive processing in language learning (Manchón & Williams, 2016; Schoonen et al., 2009), a conceptual replication study of Duong et al. (in press) with a focus on spoken tasks is necessary to offer additional insights into the potential of input- and output-based tasks for vocabulary learning in the spoken mode. Another gap in task-based research is that no evidence has been found for the effects of input-based tasks compared with input-only on vocabulary learning. To address these issues, the present study aims to compare the effects of three treatments: input-only (i.e., without tasks), spoken input-based tasks, and spoken output-based tasks on three aspects of L2 vocabulary learning: spontaneous use, form recall, and meaning recall tests. To this end, we formulated the following research questions:

1. Can vocabulary be learned from doing input-based tasks, output-based tasks, and getting input-only? If so, which aspects of word knowledge can learners acquire from input-based, output-based tasks, and input-only?
2. Is there a difference in the effectiveness of output-based tasks, input-based tasks, and input-only on vocabulary learning?

Methodology

Design

This study employed a pretest-posttest design with two groups: an experimental group and a comparison group. The experimental group completed both input-based and output-based tasks. Specifically, for the input-based tasks, the participants watched L2 videos clips and did the speaking tasks in their L1. For the output-based tasks, the participants were not exposed to the L2 video clips, but they did speaking tasks in their L2 based on pictures and L1 text prompts. Twenty target words were embedded in the tasks. The target words were counterbalanced across the input- and output-based tasks, meaning that half of the participants processed half of the target words in the input-based tasks and the other half in the output-based tasks. The other half of the participants processed the second set of the target words in the input-based tasks and the first half of the words in the output-based tasks. Participants in the comparison (i.e., input-only) group were only exposed to the target items in the L2 input without doing the tasks.

Participants

Sixty-four Vietnamese EFL university students (L1=Vietnamese, aged 18–20) participated in this experiment on a voluntary basis. Participants were from different English classes of two universities and were expected to be at the A2-B1 level, as defined by the Common European Framework of Reference for Languages (CEFR). The participants were randomly assigned to the experimental group ($n=32$) and the input-only group ($n=12$).[1] All data were collected individually. We administered a vocabulary size test to ensure that two groups were not different in their vocabulary size. The analysis showed that there was no significant difference between the two groups ($p=.08$).

Learning materials and target items

Both input- and output-based tasks were designed to elicit the target items. In other words, a task was considered successfully completed if the learners had used the target items (Loschky & Bley-Vroman, 1993). We designed all tasks on the learning management system, *Moodle*, so that the learners could complete the tasks online as face-to-face data collection with exactly the same procedure for all participants was not possible for practical reasons (e.g., participants have different studying and exam timetables). Contact between the learners and the researchers was possible via Skype during the task performance. A screen sharing tool on Skype (i.e., a tool that allows users to share their computer screens during Skype calls) was used to check the learners' task performance and guarantee the learners' integrity.

Input

The input consisted of five short L2 captioned English-language videos (total time=13 minutes). Captioned videos were chosen because they are beneficial for video comprehension (e.g., Baltova, 1999) and vocabulary learning (e.g., Montero Perez et al., 2014; Winke et al., 2010). The videos were about tourist attractions and they were taken from *Viator Travel* Youtube channel, which features videos narrated by native-speakers, about well-known tourist destinations and attrac-

---

[1]   The difference in sample size between the experimental group and the input-only group results from the fact that the timing of collecting data for the input-only group coincided with the exam period. Therefore, a considerable number of participants in the input-only group withdrew from the experiment. Data of the experimental group were collected first so data attrition of this group is far less than that of the input-only group.

tions worldwide. The lexical profile of the video was checked with Nation's Range software (Nation & Heatley, 1994), which showed that the 4,000 most frequent word families provided 95% coverage. The learners had to watch each video twice. They could pause while watching the videos, but they could not rewind the videos. By doing this, we expected the frequency of exposure in the number of viewings to be the same for all participants.

Input-based tasks

The experimental group performed two input-based tasks. In each task, the participants were asked to do the following activities in a set order: (a) read L1 emails, (b) watch L2 captioned videos (= subtitles in the L2), and (c) leave a spoken phone message in their L1. The L1 emails were used to draw the learners' attention to the target items that would appear later in the L2 videos. In these L1 emails, meaningful questions related to the target items were raised. In the reply phone messages, learners had to answer the questions from the L1 emails using information from the L2 videos. Learners had to leave the phone messages in their L1 to ensure that their focus was mainly on the L1 meaning of the target items, not on producing the L2 form. In the first input-based task, learners had to leave a message to give travel tips in Vietnamese to a friend. In the second task, they had to leave a message to suggest a tour schedule in Vietnamese to tourists. The researcher explained to the learners that the goal of the input-based tasks was: to help a friend preparing for his/her vacation trip (task 1); and to plan a tour in the imaginative role of a tour guide (task 2). Learners were allowed to look up words in a self-designed English-Vietnamese glossary (using the H5P tool), which contained words taken from the videos (both target items and others) that might be unknown to the learners (see Figure 1).

The following example illustrates how learners were prompted to attend to the meaning of a target item in the first input-based task:

Task 1
Imagine that you are a tour operator of a traveling company which specializes in operating customized tours for small groups of tourists. The situation is that you received an email, written in Vietnamese, from a tourist who has asked you to design a tour based on his requirements and to leave a phone message in Vietnamese with some information about the tour you design for him.

(This is an English translation of an excerpt of the L1/Vietnamese email that the learners received from a tourist. In the treatment, the learners read the email in Vietnamese.)

Dear Mr. Tuan,
This summer, my family is going to visit Cairns and we know from friends that your
company is one of the top companies specializing in operating customized tours for
small groups of tourists. We have some special requirements for the trip:
– *We would like to experience the panoramic view of Cairns from the sky. (elicited*
   item: hot-air balloon)
   …

In this situation, to design a tour schedule which takes into account the require-
ment of having a panoramic view of Cairns, learners have to focus on the parts
mentioning such information in the L2 video, that is, "*Join us for an insider look*
of Cairns, Australia Sailing up, up and away, in a hot-air balloon, and watching the
sun rise is one of the most incredible ways to see Cairns."

    Because the learners have to leave a phone message in their L1, they are
pushed to figure out the meaning of the English target item (i.e., hot-air balloon)
by using the given web-based glossary.

3. Key words

Click vào từ mới để xem nghĩa Tiếng Việt

∨ **cable car**

Danh từ

Cáp treo

> **bullet train**

> **delve into**

> **dolphin**

> **religious charms**

> **(to) house**

Figure 1. The web-based English-Vietnamese glossary (created by H5P Accordion tool)

Output-based tasks

Unlike in the input-based tasks, learners were not exposed to L2 captioned videos
in the output-based tasks. They performed two oral output-based tasks: (1) orally

describe your (imaginary) travel experience in English using Instagram photos as prompts, and (2) orally present in English a travel itinerary to foreign tourists based on Vietnamse tourist leaflets in which tourist information was presented in bulleted lists. We used picture prompts (i.e. Instagram photos) and L1 bulleted lists (in the leaflets) to elicit the use of target items. In addition, a web-based English-Vietnamese picture glossary was created to faciliate the learners' output production, as previous research has shown that students need to have some receptive knowledge of language forms before engaging in any output activities (Swain & Lapkin, 2007). However, the learners were not compelled to use the glossary. With such a design, the focus of these tasks is predominantly on output production, not input comprehension like in the input-based tasks. The glossary contained pictures and L2 descriptions related to the target items in the written and spoken mode (i.e. learners could see the descriptions and click on the loud speaker symbol to hear the spoken form) (see Figure 2).
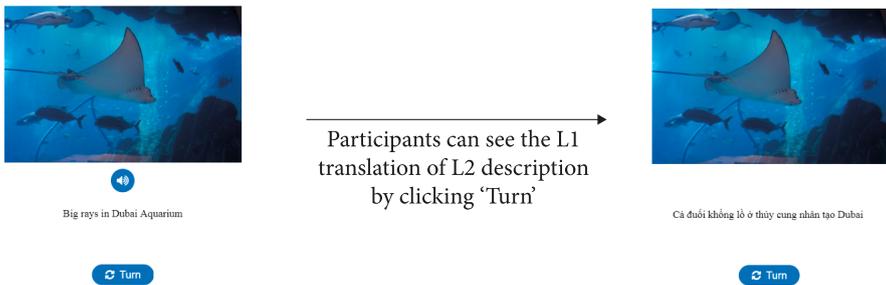

Participants can see the L1 translation of L2 description by clicking 'Turn'

Big rays in Dubai Aquarium

Cá đuối khổng lồ ở thủy cung nhân tạo Dubai

Figure 2. The web-based English-Vietnamese picture glossary (created by H5P Flashcard tool)

Target items

Twenty words (10 single words, 10 compounds) were selected as target words (for a list, see Table 1). All target items had a high level of concreteness (4–5 based on Brysbaert, Warriner and Kuperman's (2014) concreteness ratings) because pictures were used as prompts. The concreteness of compounds was checked through the concreteness level of the head noun of the compounds, viz. the word that determines the core meaning of the sequence (e.g., *balloon* in *hot-air balloon*). We used Corpus of Contemporary American English (COCA) (Davies, 2009) to check the compounds' Mutual Information (MI) scores. MI scores are typically used to measure the strength of co-occurrence of two words. Only compounds with an MI score higher than three – a commonly accepted cut-off score for an item to be considered a multi-word unit, were selected (McEnery, 2006).

Second, to guarantee that the items were not in the textbook nor taught in class in Vietnam, they were cross-checked with Vietnamese teachers. We decided to select both single words and compounds (a type of formulaic sequence) in our target items given that most task-based research has focused on single words.

The frequency of encounters with the target items in the videos was controlled: all items appeared only once, and video rewinding was not allowed. To avoid any confounding effect from the target items, the items were counterbalanced across the input- and output-based tasks.

Table 1. List of target items with their values

Single words	Frequency range	Concreteness	Compounds	MI Score	Concreteness
Statue	7277	4.93	Bullet train	13.98	4.79
Submarine	2582	4.8	Cable car	12.01	4.92
Aboriginal	2471	3.12	Observation wheel	15.10	4.86
Spear	2601	5	Glass bottom boat	12.32	4.93
Dolphin	1889	4.96	Hot air balloon	10.93	4.96
Armor	4720	4.76	Northern light	14.00	4.21
Sword	7911	4.93	Hot spring	12.97	3.89
Mosque	4229	4.54	Zip line	11.93	4.5
Ray	22139	3.57	Observation deck	15.26	4.77
Peacock	1385	5	Chair lift	13.59	4.68

Test instruments

Vocabulary size test

To control for individual differences in overall prior vocabulary knowledge in English, a vocabulary size test was administered to the participants, viz. the bilingual version of Nguyen and Nation's (2011) Vocabulary Size Test (VST). We used the bilingual version to reduce the effects of non-vocabulary factors (e.g., grammar, reading skill) influencing the test result. The test is a frequency-based meaning recognition test, containing 140 items: 10 items from 14 frequency bands of 1000 words (1K-14K). A pilot showed that the test procedure was too long, so we decided to develop a shorter version (70-item) of the bilingual VST in the same way as Beglar (2010) did for the monolingual version: five items per frequency

level were randomly selected from the 140-item test. The 70-item version pro-
duced reliable test scores in terms of internal consistency (Cronbach's alpha = .82,
$n = 44$).

Vocabulary knowledge tests

Three tests were administered to measure learners' vocabulary learning gains: a
test focusing on spontaneous use of the target items in an oral narrative, an oral
form recall test, and an oral meaning recall test. To control for test effects, we
added 15 distractors to the form recall and the meaning recall tests. The distractors
consisted of 8 single nouns and 7 nominal compounds (for the list, see Table 2).
The distractors were selected on the basis of the same criteria as the target items.
To avoid the risk that earlier tests can influence test scores on later tests, the three
vocabulary tests were carefully sequenced: learners first took the spontaneous use
test, then the form recall and finally the meaning recall test. The same test battery
was used for both pre- and post-test but the items in the tests were ordered differ-
ently. To control for a test effect, learners had to do an unrelated test between the
form recall and the meaning recall test, viz. a 10-question listening comprehen-
sion test.

Table 2. List of distractor items with their values

Single words	Frequency range	Concreteness	Compounds	MI Score	Concreteness
Asparagus	2680	4.96	Smoke detector	9.89	4.8
Broom	2253	4.89	Conveyor belt	14.10	4.9
Canon	4293	4.56	Drainage ditch	12.98	4.8
Cashew	2776	4.92	Spice rack	10.08	4.8
Cradle	2282	4.93	Ski lodge	8.76	4.79
Faucet	1384	4.48	Oil refinery	9.86	4.79
Freight	2526	4.21	Flea market	10.77	4.62
			Rocking chair	11.33	4.93

a. Oral spontaneous use test: In this test, participants were asked to describe
    their imaginary travel experience to five places (Singapore, Tokyo, Cairns,
    Iceland, Dubai) based on 20 pictures, corresponding to 20 target items. The
    participants had 3 minutes for preparation and 3 minutes at most to talk.
    The time of 3 minutes was chosen because the pilot testing showed that such
    amount of time was sufficient for learners to prepare and to talk.

      Instruction: Imagine that you have just come back to Vietnam from a trip to five famous destinations (Singapore, Tokyo, Cairns in Australia, Iceland, and Dubai) and you took twenty 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 3 minutes at most to talk and 3 minutes for preparation. You are not allowed to use a dictionary or any other reference sources.

b.   Oral form recall test: In this test, participants had to say the word that best describes a given picture. The first letter of the item was given as a clue. The L1 meaning of the item was not given.

     Example: We skied together towards the c_ _ _ _ l_ _ _ (2 words) that would take us to the mountain.

c.   The oral meaning recall test was an oral English-Vietnamese translation test. The items were presented in L2 written and spoken form because learners were exposed to both written and spoken form through captioned videos (in the input-based tasks), and through the web-based picture dictionary (in the output-based tasks). The participants had to say the Vietnamese meaning or explanation of the target items without context.

     Example: Glass bottom boat   _____

We administered the tests to all participants individually via Skype. The tests were designed on PowerPoint and were run on the researcher's computer. The test items appeared one by one. The screen sharing tool on Skype was used to share the screen of the researcher's computer, thus the learners could see the tests and say the answers via the Skype voice chat feature. Learners' webcams were activated to check the learners' integrity while doing the tests.

Procedure

There were five sessions (see Figure 3). In the first session, both the experimental and input-only groups took the Vocabulary Size Test in class. One week later in session 2, the experimental group worked on a task familiarity session in class. After one week, participants of the experimental and input-only groups did the pretests. In the fourth week, the experimental group did the tasks while the input-only group was exposed to audiovisual input only (= watching the L2 captioned videos) and the English-Vietnamese web-based glossary without doing the tasks. One week later, both groups took unannounced posttests. The participants had

three hours in total to finish the learning tasks, and sixty minutes to do the pretests/posttests.
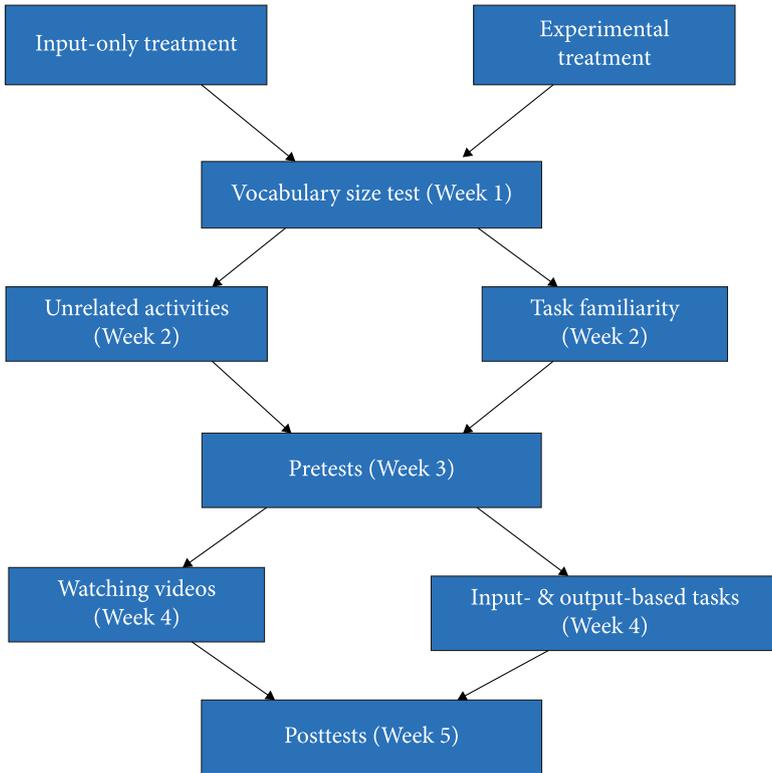

Figure 3. Research procedure

Scoring and analysis[2]

The maximum score for the spontaneous use test was 20 (1 point per target item) and the maximum score for the form recall and meaning recall tests was 35 points for 20 target items and 15 distractors. The tests were scored dichotomously. Learn-

---

2. To explore the role of prior vocabulary knowledge and lexical type in the vocabulary learning gains from input-, output-based tasks, and input-only, we also ran a secondary analysis (i.e., logistic regression) with the treatment parameter (input tasks vs. outputs task vs input-only) and the lexical type parameter (single word vs. compound) as the main predictors and prior vocabulary knowledge as a covariate for the form recall and the meaning recall posttests. The analyses yielded the same findings for Research Question 2 (see Appendix).

ers received 0 for an incorrect response and 1 for a fully correct response. The first author (a Vietnamese-English bilingual) first scored all the responses. To check reliability, two other experienced Vietnamese teachers of English independently scored 10% of all tests which were selected randomly. No differences were found between the three assessors' scores in the spontaneous use and form recall tests. The interrater reliability for the meaning recall test was $r=.98$ in the pretest and in the posttest.

The results were analyzed with SPSS 25. We used non-parametric tests to analyze the data as the assumption of normality was violated. Specifically, Wilcoxon-Signed Rank tests were conducted to determine if there were significantly more learning gains for target items than for distractors and if the learning could be attributed to the treatment (research question 1). Because we only tested the distractors in the form recall and meaning recall tests, we only ran the Wilcoxon-Signed Rank analyses for these two tests. The learning gains of the spontaneous use test were analyzed based on the descriptive statistics but no further analysis will be conducted on this test.

To answer research question 2, firstly, we ran the Wilcoxon-Signed Rank tests to analyze the differential effect between the input- and output-based tasks as participants of the experimental group completed both input- and output-based tasks. Then, the Mann Whitney U tests were run to analyze the difference in learning gains between the input-based tasks and the input-only as well as between the output-based tasks and the input-only. The Mann Whitney U tests were computed with the absolute gains calculated at the item level (= posttest score – pretest score per item) rather than by subtracting the total pretest score from the total posttest score. A separate analysis was computed per posttest.

Results

Prior vocabulary knowledge

Table 3 describes the descriptive results of the vocabulary size test. An independent samples *t*-test showed that even though the input-only group obtained higher scores than the experimental group, there was no significant difference between the experimental and input-only group in terms of prior vocabulary knowledge, with $t=-1.90$, $p=.078$, $df=13.74$, $d=2.69$. Given that two groups did not have homogenous variance ($p=.003$), we used the *t*-test results in the case of unequal variance.

Table 3. Mean score and estimate in word family per group

	Min	Max	Mean score (Max=70) (SD)	Confidence interval
Experimental (*N*=32)	21	48	33.53 (1.07)	31.34–35.72
Input-only (*N*=12)	27	59	39.75 (3.09)	32.95–46.55

Learners' responses to target items during input- and output-based tasks performance

We checked whether learners could comprehend and produce the target items during task performance. A response was considered accurate if the learners said the Vietnamese meaning of target items in the input-based tasks or if they said the correct English form of target items in the output-based tasks. A few learners used some target items in the L2 instead of the L1 (i.e., codeswitching) while performing the input-based tasks but they immediately switched to L1 when being reminded of the input-based task instruction. Mean scores of the correct responses on each task type were compared (see Table 4) using a Wilcoxon Sign-Rank Test because the assumption of normal distribution was violated. The analysis showed no significant difference in learners' responses to target items in the two task types (*p*=.433), which implies that the focus on form techniques used in two task types had similar effects. The descriptives in Table 2 show that we were successful in eliciting learners' processing of the target items in the two task types. Further, there was no difference between the two task types in terms of correct use of the target items.

Table 4. Learners' responses to target items in input-and output-based tasks

	Minimum	Maximum	Mean (Max=10)	SD
Input-based tasks	6	10	9.31	.176
Output-based tasks	5	10	9.19	.235

Research question 1: Can words be learned from doing input-based tasks, output-based tasks, and input-only with no tasks?

The descriptive results of the three vocabulary tests (use, form recall, meaning recall) are reported in Table 5. The maximum score possible was 35 for the meaning recall, 35 for the form recall tests (20 for the target items, 15 for the distractors), and 20 for the spontaneous use test (20 for the target items only). Learning gains were found for all three levels of vocabulary knowledge. Scores on the form recall test were lower than on the meaning recall test, but the absolute learning gains

(i.e. the difference between the pre- and post-test scores) were almost the same for the form recall and meaning recall aspect. We also found learning gains for the distractors in the form recall and meaning recall tests possibly due a test effect.

Table 5. Learning gains – target item and distractors

	Pre-test M (SD)	95% CI	Post-test M (SD)	95% CI	Abs. gains M (SD)	95% CI
Target items (max = 20)						
Use	2.00 (0.43)	1.11–2.89	6.09 (0.68)	4.70–7.48	4.13 (0.48)	3.13–5.12
Form recall	4.06 (0.44)	3.17–4.96	9.16 (0.73)	7.65–10.66	5.13 (0.52)	4.05–6.20
Meaning recall	6.88 (0.60)	5.64–8.11	12.50 (0.7)	11.08–13.9	5.78 (0.47)	4.80–6.76
Distractors (max = 15)						
Form recall	1.19 (0.17)	0.83–1.55	1.84 (0.22)	1.39–2.29	0.84 (0.16)	0.50–1.19
Meaning recall	2.19 (0.27)	1.62–2.76	3.31 (0.35)	2.58–4.04	1.31 (0.31)	0.66–1.96

Note.
CI = Confidence interval

The Wilcoxon-Signed Rank test showed that there was a significant difference in learning gains between the target items and the distractors. The participants learned more target items than the distractors in the form recall test ($Z = -5.31$, $p < .000$), and the meaning recall test ($Z = -5.27$, $p < .001$). The results indicate that the learning gains could therefore be ascribed primarily to the task effects rather than the test effect. However, given that there was no distractor in the spontaneous use test, it remains uncertain whether the learning gains in this test can be attributed to the tasks. Therefore, no further analysis was conducted for the spontaneous use test.

Research question 2: Is there a difference in the effectiveness of input-based tasks, output-based tasks, and input-only on vocabulary learning?

The descriptive statistics (see Table 6) revealed that there were gains from pretests to posttests in the input-only condition, input-based tasks as well as in the output-based tasks. Table 6 also showed that words were learned slightly better in the

input- and output-based tasks than in the input-only condition at the form recall as well as the meaning recall level. The Wilcoxon-Signed Rank tests showed that there was no significant difference in gains between the input- and output-based tasks in the form recall test ($Z=-.340$, $p=.734$) and in the meaning recall test ($Z=-.041$, $p=.967$). However, the Mann-Whitney U tests revealed significant differences in learning gains between the between the input-based tasks and the input-only in the form recall test ($Z=-3.681$, $p<.001$) and the meaning recall test ($Z=-3.524$, $p<.001$). Similarly, the Mann-Whitney U tests also displayed significant differences in gains between the output-based tasks and the input-only in the form recall ($Z=-3.439$, $p=.001$) and the meaning recall test ($Z=-3.909$, $p<.001$).

Discussion

The present study extends previous research on input-based tasks by exploring another type of input-based task than listen-and-do tasks and by comparing the effects of input-based with output-based tasks as well as with input only on vocabulary learning. The findings suggest that input-based tasks, output-based tasks, and input-only condition can all facilitate L2 vocabulary learning, but to different degrees. Vocabulary was learned better in the input-based tasks than in the input-only condition, whereas there were no differences in learning gains between the input- and output-based tasks.

Effects of input- only on L2 vocabulary learning in comparison to input-based and output-based tasks

The study suggests that viewing L2 captioned videos with a glossary available resulted in learning gains in all three tests: spontaneous use, form recall and meaning recall. The findings, thus, contribute to previous research which has shown that incidental vocabulary acquisition through viewing captioned audiovisual input does occur (Montero Perez et al., 2014; Peters, 2019; Peters et al., 2016). The gains; however, were quite modest. The input-only group learned on average 1 to 2 words in three tests (i.e. use, form recall, meaning recall) after watching the videos, whereas in Nguyen and Boers' (2018) study, learners learned approximately 3 words at the meaning recall level 2 weeks after watching a TED Talk. However, participants in their study were encouraged to take notes between two times of viewing and they had to answer several yes/no content questions. In our study, note-taking was not permitted and no input-based exercise was given, which might have played a role in limiting the learners' working memory and led to lower learning gains. Additionally, research has shown that frequency of occur-

Table 6. Learning gains in input-based tasks, output-based tasks, and input-only with no tasks

	Use						Form recall						Meaning recall					
	Pretest		Posttest		Abs. gains		Pretest		Posttest		Abs. gains		Pretest		Posttest		Abs. gains	
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
Input-only	2.16 (0.77)	0.45–3.87	3.00 (0.73)	1.37–4.62	1.25 (0.57)	0.02–2.52	4.83 (1.01)	2.58–7.08	6.50 (1.13)	4.00–8.99	1.91 (0.48)	0.85–2.98	7.66 (1.36)	4.67–10.66	9.00 (1.41)	5.88–12.1	1.75 (0.60)	0.41–3.08
Input-based tasks	1.09 (0.23)	0.62–1.56	2.97 (0.38)	2.19–3.74	1.88 (0.27)	1.32–2.43	2.13 (0.22)	1.66–2.59	4.63 (0.39)	3.82–5.43	2.50 (0.31)	1.87–3.13	3.56 (0.40)	2.74–4.39	6.34 (0.39)	5.54–7.15	2.78 (0.32)	2.11–3.45
Output-based tasks	0.91 (0.22)	0.45–1.37	3.13 (0.35)	2.40–3.85	2.25 (.028)	1.67–2.83	1.88 (0.28)	1.29–2.46	4.50 (0.38)	3.71–5.29	2.66 (0.29)	2.06–3.25	3.31 (0.38)	2.54–4.09	5.91 (0.39)	5.11–6.70	2.75 (0.26)	2.21–3.29

Note.
Max score possible in three tests for the input-only group = 20, Max score possible in three tests for items learned input-based tasks = 10 and items learned in output-based tasks = 10

rence is positively related to word learning (Peters & Webb, 2018; Peters, 2019). In our study, each target item occurred only once in the video, which might be another reason for the low learning gains.

Our study showed that learners who did the input-based tasks could learn significantly more words than the input-only group, who only watched the video clips. Our findings, thus, seem to support previous research that input-plus activities could lead to vocabulary gains superior to those of input-only (e.g., Hill & Laufer, 2003; Mason & Krashen, 2004; Peters et al., 2009). Yet, given that the experimental group spent more time on the input-based tasks than the input-only group spent on watching the videos, the learning difference could also be attributed to the difference in time-on-task. The findings should therefore be interpreted with care.

L2 vocabulary learning from input-based tasks and output-based tasks

First, our study reveals that input- and output-based tasks can develop vocabulary knowledge at the level of spontaneous use of words in an oral narrative with picture prompts, oral form recall and oral meaning recall. On average, the experimental group learned 5 out of 16 items at the form recall level and 6 out of 13 items at the meaning recall level. This amount of learning is encouraging, especially since learning was measured one week after the task performance. Our results confirm Duong et al.'s (in press) findings which used the same tasks, test formats and target items, but focused on written tasks and written vocabulary knowledge.

The present study used spoken tasks and tested learning gains in oral vocabulary tests (with an exception of the meaning recall test which tested the comprehension of both written and spoken form). It should be noted that the gains were slightly higher in Duong et al. (in press) compared to this study, that is, two items more for each test (spontaneous use, form recall, and meaning recall). Given that we used the same target items, a similar test format and comparable participants' profiles, we assume that task modality (speaking vs. writing) might have played a role, and writing tasks seem to facilitate vocabulary acquisition better than speaking tasks. Yet, given that our study does not directly compare the two task modalites, further experimental research which directly compares the two modes is encouraged.

Second, we found that even though spoken input-based tasks resulted in more gains for meaning recall, and spoken output-based tasks led to larger gains for form recall, the difference between two task types for each aspect was not significant. Unlike Duong et al. (in press), this finding does not offer evidence for the Transfer Appropriate Processing (TAP) theory (Morris et al., 1977). However, our finding was consistent with Shintani (2011) who found that PPP activi-

ties and input-based tasks had similar effects on vocabulary learning at the level of meaning recall and form recall. She argued that the opportunities for interaction with the teacher, in the form of clarification requests/ confirmation checks in the listen-and-do tasks, might have enabled the learners to comprehend the input as well as use L2 forms of the target items. Therefore, learners who performed the listen-and-do tasks were assumed to have almost the same opportunities as the PPP groups to process the meaning and the L2 form. In our study, while learner-researcher interaction in the input-based tasks was not encouraged, the qualitative analysis of the learners' responses during task performance shows that sometimes learners naturally used the target items in the L2 in their L1 phone messages. This phenomenon of codeswitching might partially explain learners' significant gains in the form recall test.

It should be noted that there are a number of limitations. A first limitation concerns the small number of target items and the short period of instruction. It is worth examining whether similar effects might be obtained in conceptual replications with other lexical types and longer times of instruction. In addition, a factor that might influence the task effects is learners' use of strategies when watching the video (e.g., pausing or freeze framing) which might influence the level of noticing to target items. Further studies might explore how these strategies moderate the effects of captioned videos on language acquisition. Also, we had a fairly small numbers of participants in the input-only group compared with the experimental group, which limits the generalizability of our findings. Another limitation pertains to the meaning recall test. As the items were presented in both L2 written and spoken form in the meaning recall test, it seemed difficult to know whether the learners could recall the spoken form or the written form of the target items. Thus, in future studies, the use of meaning recall tests in which items are presented only in L2 spoken or written form would be an asset.

Conclusion

This study has investigated the comparative effects of input-only, spoken input-based tasks, and spoken output-based tasks on vocabulary acquisition. Despite the fact that input- and output-based tasks were operationalized differently, both types of tasks led to similar gains in comprehension (= meaning recall of written and spoken form) and productive vocabulary knowledge (= spontaneous use of words in oral narratives with picture prompts, and oral form recall). In addition, the study revealed that input-based tasks led to significantly more learning gains than input-only, which indicates that vocabulary learning can be enhanced if words are processed again in tasks.

In terms of pedagogical implications, together with previous research (Duong et al., in press; Shintani, 2012), our study suggests that input-based tasks can foster vocabulary learning; and we recommend that teachers might consider employing input-based tasks in written as well as oral modality to teach vocabulary to A2-B1 level students. In addition, this study again emphasizes the value of input-plus activities in vocabulary teaching. It has been well established that TV viewing is useful for incidental vocabulary learning. However, important words might not be learned incidentally and thus follow-up activities such as input/output-based tasks might be administered to focus learners' attention to specific words.

References

Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *Canadian Modern Language Review*, *56*, 32–48. https://doi.org/10.3138/cmlr.56.1.31

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*(1),101–118. https://doi.org/10.1177/0265532209340194

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Davies, M. (2009). The 385+million word corpus of contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*, 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

Duong, P. T., Montero Perez, M., Desmet, P., & Peters, E. (in press). Differential effects of input-based and output-based tasks on L2 vocabulary learning. *Canadian Journal of Applied Linguistics*.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, *44*, 449–491. https://doi.org/10.1111/j.1467-1770.1994.tb01114.x

Erlam, R., & R. Ellis. (2018). Input-based tasks for beginner-level learners: An approximate replication and extension of Erlam & Ellis (2018). *Language Teaching*, 1–22. https://doi.org/10.1017/S0261444818000216

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford University Press.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching*, *41*(2), 87–106. https://doi.org/10.1515/iral.2003.007

Loschky, L. & R. Bley-Vroman. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Multilingual Matters.

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

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *Modern Language Journal*, *92*(1), 114–130. https://doi.org/10.1111/j.1540-4781.2008.00690.x

Manchón, R., & Williams, J. (2016). L2 writing and SLA studies. In R. Manchón & P. Matsuda (Eds.), *The handbook of second language writing* (pp. 567–586). De Gruyter Mouton. https://doi.org/10.1515/9781614511335-029

Mason, B., & Krashen, S. (2004). Is form-focused vocabulary instruction worth-while? *RELC ournal*, *35*(2), 179–185. https://doi.org/10.1177/003368820403500206

McEnery, T. (2006). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Routledge.

Min, H-T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, *58*(1), 73–115. https://doi.org/10.1111/j.1467-9922.2007.00435.x

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118–141.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nation, I. S. P., & Heatley, A. (1994). Range: A program for the analysis of vocabulary in texts [Computer software]. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources /vocabulary-analysis-programs

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, *14*(4), 397–419. https://doi.org/10.1177/1362168810375364

Nguyen, L. T. C., & Nation, I. S. 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*. https://doi.org/10.1002/tesq.441

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. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audivisual input. *TESOL Quarterly*, *53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., Heynen, E., Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, *59*(1), 113–151. https://doi.org/10.1111/j.1467-9922.2009.00502.x

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, *36*, 73–97. https://doi.org/10.1017/S0267190516000015

Révész, A. (2017). Replication in task-based language teaching research: Kim (2012) and Shintani (2012). *Language Teaching*, *52*(3), 374–384. https://doi.org/10.1017/S0261444817000283

Schoonen, R., Snellings, P., Stevenson, M., & van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R.M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 77–101). Multilingual Matters. https://doi.org/10.21832/9781847691859-007

Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research*, *15*(2), 137–158. https://doi.org/10.1177/1362168810388692

Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research*, *16*(2), 253–279. https://doi.org/10.1177/1362168811431378

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginning-level learners. *TESOL Quarterly*, *47*(1), 36–62. https://doi.org/10.1002/tesq.54

Shintani, N. (2015). The incidental grammar acquisition in focus on form and focus on forms instruction for young, beginner learners. *TESOL Quarterly*, *49*(1), 115–140. https://doi.org/10.1002/tesq.166

Swain, M. (1998). Focus on form through conscious reflection. In Doughty, C., & Williams, J. (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 471–484). Lawrence Erlbaum Associates.

Swain, M., & Lapkin, S. (2007). 'Oh, I get it now!' From production to comprehension in second language learning. In D.M. Brinton & O. Kagan (Eds.), *Heritage language acquisition: A new field emerging* (pp. 301–319). Lawrence Erlbaum Associates.

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

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

Rodgers, M., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics*, *171*(2), 191–220. https://doi.org/10.1075/itl.18034.rod

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86. http://llt.msu.edu/vol14num1/winkegasssydorenko.pdf

Appendix.   GEE analyses for form recall and meaning recall test

We used GEE to fit a repeated measured logistic regression as GEE does not assume normality and homogeneity of variance and the technique allows us to include observations per item per participant. The dependent variable in the models (one model for the form recall test, and one for the meaning recall test) was the scores on the posttests of target items that were not correctly

answered in the pretests. The following predictors were entered as main effects in the model: 'treatment' (input-only vs. input-based tasks vs. output-based tasks), and 'lexical type' (single vs. compound). The learners' vocabulary size was entered as a covariate. For all models, predictors that did not contribute significantly to the models were then removed one by one, until the final models contained only predictors with a p-value lower than .05.

For the form recall test, the GEE analysis (see Table 1) revealed that the treatment parameter (input-only vs. input-based tasks) ($p=.002$) significantly contributed to the model. The odds ratio showed that if the items were processed in the input-based tasks, learners were almost 3 times (1/.328 = 3.04) more likely to produce such items correctly in the form recall posttest. However, the possibility that learners produced the items correctly was almost the same if the items were processed in the input-based tasks and the output-based tasks ($p=.709$). A Bonferroni post-hoc test showed that the odds that learners produced the target items correctly differed when the items processed in the output-based tasks and in the input-only condition ($p=.002$). Additionally, while vocabulary size significantly predicted the learning gains ($p=.022$), 'lexical type' did not.

Table A.1 GEE analysis for the form recall posttest

Parameter	Wald chi-square	df	Sig	B	Exp(B)	CI	
						Lower	Upper
Intercept	11.00	1	.001	−1.88	.152	.050	.463
Input-only = 2	9.60	1	.002	−1.11	.328	.162	.664
Output task = 1	.139	1	.709	−.056	.946	.706	1.26
Input task = 0	0						
Vocabulary size	5.274	1	.022	.000	1.000	1.000	1.000

Note.
Exp(B) = odds ratio, CI = confidence interval

For the meaning recall test, the GEE analysis showed significant effects of one predictor: the treatment parameter (input-based tasks vs. input only) ($p=.018$) (see Table 2). Learners were almost 2.5 times (1/.404 = 2.47) more likely to say the correct meaning of the target items in the posttest if the items were processed in the input-based tasks. Similar to the findings in the form recall test, the possibility that learners produced the items correctly was not significantly different when the items were processed in the input-based tasks and the output-based tasks ($p=.654$). A Bonferroni post-hoc test showed that that the odds that learners produced the items correctly was significantly different if the items were processed in the output-based tasks and input-only condition ($p=.006$). We did not find significant effects of two predictors, 'lexical types' and 'vocabulary size'.

Table A.2 GEE analysis for the meaning recall posttest

Parameter	Wald chi-square	*df*	Sig	B	Exp(B)	CI Lower	CI Upper
Intercept	11.16	1	.001	−.593	.553	.391	.783
Input only = 2	5.59	1	.018	−.907	.404	.190	.856
Output task = 1	.201	1	.654	.103	1.10	.707	1.73
Input task = 0	0						

Note.
Exp(B) = odds ratio, CI = confidence interval

Address for correspondence

Phuong Thao Duong
Department of Linguistics
KU Leuven
Sint-Andriesstraat 2
2000 Antwerpen
Belgium
phuongthao.duong@kuleuven.be

Biographical notes

Phuong Thao Duong is a doctoral student in the Department of Linguistics at KU Leuven. She is also affiliated with the Itec-imec group at KU Leuven. Her research interests include task-based language learning, vocabulary learning in a foreign language, and technology-supported language teaching/learning.

Maribel Montero Perez (https://research.flw.ugent.be/en/maribel.monteroperez) is an assistant professor of second language acquisition and language didactics at Ghent University (Belgium). Her research interests include, amongst others, multimodal input for L2 learning, technology- enhanced listening, and vocabulary learning.
https://orcid.org/0000-0002-0868-588X

Piet Desmet is Full Professor in French & Applied Linguistics at KU Leuven (Belgium) and its KULAK Kortrijk Campus. He leads the KU Leuven & imec research team ITEC devoted to impactful digital innovation in education and related fields and is scientific director of the imec smart education research program. He is also the editorial board member of the journal ITL-International Journal of Applied Linguistics and CALICO Journal.
https://orcid.org/0000-0002-9849-0874

Elke Peters is an associate professor at KU Leuven, where she coordinates the research group "Language, education and society". Her main research interests involve incidental and deliberate vocabulary learning in a foreign language. She is interested in how different types of input can contribute to vocabulary learning. Her recent research also focuses on the potential of out-of-school language learning. She has published her research in Language Learning, Studies in Second Language Acquisition, Language Teaching Research, and TESOL Quarterly. She is also the editor of the journal ITL-International Journal of Applied Linguistics.

https://orcid.org/0000-0001-7273-3850

Publication history