

Differential Effects of Input-based and Output-based Tasks on L2 Vocabulary Learning

Phuong-Thao Duong
KU Leuven, imec research group itec

Maribel Montero Perez
Ghent University

Piet Desmet
KU Leuven, imec research group itec

Elke Peters
KU Leuven, Department of Linguistics

Abstract

This paper reports on a quasi-experimental study that explored the differential effects of input- and output-based tasks on L2 vocabulary learning. The study adopted a pretest-posttest within-participants design, with sixty Vietnamese EFL university students. Participants in the experimental group completed four input and four output tasks in an online learning environment. The target items (ten single words and ten compounds) were counterbalanced across the tasks. The study measured vocabulary gains at four levels of sensitivity: spontaneous use of the target items, form recall, meaning recall and meaning recognition. The analyses showed that the input-based tasks resulted in higher vocabulary gains at the meaning recall level, whereas the output-based tasks resulted in better scores at the form recall level. No difference was found in the spontaneous use and meaning recognition tests.

Résumé

Cet article rend compte d'une étude quasi expérimentale qui a exploré les effets différentiels des tâches basées sur les entrées et les sorties sur l'apprentissage du vocabulaire L2. L'étude a adopté une conception intra-participants prétest-post-test, avec soixante étudiants universitaires vietnamiens EFL. Les participants du groupe expérimental ont effectué quatre tâches d'entrée et quatre tâches de sortie dans un environnement d'apprentissage en ligne. Les éléments cibles (dix mots simples et dix composés) ont été contrebalancés entre les tâches. L'étude a mesuré les gains de vocabulaire à quatre niveaux de sensibilité : utilisation spontanée des éléments cibles, rappel de forme, rappel de sens et reconnaissance de sens. Les analyses ont montré que les tâches basées sur les entrées entraînaient des gains de vocabulaire plus élevés au niveau du rappel de sens, tandis que les tâches basées sur les sorties entraînaient de meilleurs scores au niveau du rappel de forme. Aucune différence n'a été trouvée dans les tests d'utilisation spontanée et de reconnaissance de sens.

Differential effects of Input-based and Output-based Tasks on L2 Vocabulary Learning

Second language acquisition (SLA) research has demonstrated that learning gains from entirely meaning-focused instruction are often modest. Thus, researchers have proposed that learners' attention should be directed to target linguistic elements in a meaning-centred activity to improve language acquisition (Schmidt, 2001). This can be done through focused tasks (Ellis et al., 2002). So far, research on focused tasks has been mostly concerned with grammar, not vocabulary learning (Keck et al., 2006), even though there is evidence that form-focused instruction is also essential for vocabulary (Laufer, 2005). In addition, the effectiveness of focused tasks has been compared to entirely meaning-centred activities (e.g., de la Fuente, 2002) or Focus-on-Forms activities which focus on discrete, isolated, specific language forms rather than meaningful activities (e.g., Erlam & Ellis, 2018; Shintani, 2013, 2015). Yet, studies in which different types of focused tasks have been compared are relatively scarce. Input-based and output-based tasks in particular have not been investigated with the same level of attention. Despite a large number of studies on output-based tasks (85 studies from 2006 to 2016 according to Plonsky and Kim's (2016) meta-analysis), only a few studies have focused on input-based tasks. To fill those gaps, the current study investigates the differential effects of input-based and output-based tasks on vocabulary learning.

Background

Focused Tasks

In task-based teaching, tasks are generally defined as activities that satisfy four main criteria: meaning is primary, learners' linguistic resources are not restricted, there is some kind of information gap between interlocutors, and the linguistic outcome is not the only task outcome (Ellis, 2009). A focused task is a type of task within the *focus on form* approach which aims "to elicit the use of specific linguistic features in the context of meaning-centred language use" (Ellis et al., 2002, p.420). The term '*form*' in the definition refers to formal linguistic aspects such as phonology, vocabulary, grammar, or pragmatics (Ellis, 2009). In the case of focused tasks, the 'form' is usually pre-determined. Thus, focused tasks are classified as a kind of '*planned focus on form*' (Ellis et al., 2002). Focused tasks have been related mainly to the teaching of grammar (see the review by Ellis, 2003). Laufer (2005), however, argued that focused tasks can also be used for vocabulary teaching and showed that focused tasks are an effective method for teaching vocabulary. In the next sections, we will focus on two types of focused tasks and their roles in L2 vocabulary learning

Output-Based Tasks

Swain's (1998, 2005) Output Hypothesis argues that output does not just provide opportunities for language use, but might aid SLA in many ways: promoting noticing, providing learners with opportunities to test hypotheses about how the target language works, and chances to reflect on their own language use. Numerous studies have demonstrated that the noticing function of output is a mechanism that prompts L2 learners to become aware of the gaps in their existing interlanguage and the target language they need to learn in order to express their ideas (e.g., Izumi, 2002; Izumi & Bigalow, 2000; Swain & Lapkin, 1995; Van den Branden, 1997). Izumi (2003) elaborated on the psycholinguistic mechanism that underlines the noticing function and argued that the nature of output activities tends to arouse a series of cognitive processes such as "lexical search and retrieval grammatical encoding, syntactic building" (p.182). During such cognitive processes, learners would be naturally pushed to become conscious of the problems in their interlanguage capabilities; therefore, they might pay close attention to linguistic features in the subsequent input, if any, to fill their gaps. Izumi and colleagues demonstrated that such focused attention was useful for L2 language acquisition (Izumi, 2002; Izumi & Bigelow, 2000).

On the postulate that output promotes noticing and might lead to L2 acquisition, a wealth of task-based studies has examined the role of output-based tasks (i.e., tasks in which learners are engaged in producing meaningful spoken or written output) in L2 learning. However, compared to the large number of studies on grammar (for a review, see Plonsky & Kim, 2016), only a few studies have investigated word learning. These studies often focused on the comparison of collaborative and individual output tasks (e.g., Kim, 2008; Nassaji & Tian, 2015). A few studies have explored the effects of output-based tasks in comparison with non-output tasks/activities; however, their focus has been mainly on grammar (e.g., Révész, 2007; Révész & Han, 2006). Only recently, research has started to explore how output tasks affect vocabulary learning. For instance, in Nguyen and Boers' (2018) study, all participants watched an L2 TED Talk twice, but the experimental group completed an oral summary in L2 between two times of video viewing whereas the control group did not. Slightly more gains (i.e. meaning recall) were found in the experimental group than in the control group, which indicated the effectiveness of the oral summary task in enhancing vocabulary acquisition from viewing a TED Talk. The authors argued that the output task helped learners to recognize the problematicity in their language use and prompted them to process the input (i.e., TED Talk) with a clearer purpose in mind, which might have led to the acquisition of target items.

Input-Based Tasks

An input-based task is conceptualized as a type of focused task in which learners process input via listening or reading; L2 production from learners is not required but not prohibited (Ellis, 2009). An input-based task is typically designed with two purposes: a) to engage learners in input comprehension, and b) to attract learners' attention to specific linguistic features in a meaningful context. Eye-tracking studies have shown that learners have longer fixations for new words than known words in written input, and this increased attention to new words contributes to the learning gains. (Godfroid et al., 2018; Pellicer-

Sánchez, 2016). Input-based tasks have been typically operationalized as *listen-and-do* tasks, that is, tasks where learners have to listen to verbal input and demonstrate their comprehension of target language non-verbally, e.g., choosing the correct picture related to the target L2 forms between two options. It is assumed that, with such a design, learners can naturally attend to the target L2 forms in order to complete the task. Given that production is not prohibited, learners are free to engage in interaction with the teacher and learn the target words from the interaction.

Much of the research on listen-and-do tasks has been conducted to investigate the effect of different types of input and output on vocabulary acquisition (e.g., Ellis et al., 1994; Ellis & He, 1999; de la Fuente, 2002). Recently, Shintani (2012) has begun to explore the effects of listen-and-do tasks on vocabulary learning more thoroughly with Japanese EFL beginner-level learners (aged 6-8). The results showed that input-based tasks led to significant vocabulary gains. Erlam and Ellis (2018) conducted an experimental study into the effects of listen-and-do tasks for French ESL beginner-level learners (aged approximately 13). Their findings mirrored Shintani's (2012), viz. L2 French learners were successful in improving vocabulary knowledge at the level of form-meaning association and form recall. Noticeably, based on a qualitative analysis of learners' interactions while performing input-based tasks, Shintani (2012) suggested that the processes of meaning negotiation might have enabled learners to comprehend the input as well as notice new or partially learned L2 forms. However, it should be noticed that Shintani's studies focused on learners at beginner-level and listen-and-do tasks. Thus, further investigation into other types of input-based tasks and learners at higher proficiency levels is warranted to determine whether similar findings can be found, as suggested by Révész (2017).

Input-Based vs. Output-Based Instruction

In task-based research, studies that have compared the effects of input- and output-based tasks on vocabulary learning are scarce. Of potential relevance might be the studies by Shintani (2011, 2013) who compared the effects of input-based tasks and PPP activities (i.e., activities in which language is treated as an object for learning, not a tool for communication). Shintani found that both input-based tasks and PPP activities improved learners' receptive and productive knowledge of nouns (Shintani, 2011), but input-based tasks led to higher productive knowledge of adjectives than PPP (Shintani, 2013). She explained these findings in terms of learners' deeper processing of the target items, being the result of learner-initiated interactions during the input-based task performances, such as confirmation checks and clarification requests. The question is whether similar effects could be found if input-based tasks were compared with output-based tasks instead of PPP activities.

Of relevance might be a few vocabulary studies that have compared receptive learning (i.e., memorizing the L1 meaning of words in a word list) and productive learning (i.e., memorizing the L2 form of words in a word list) on vocabulary acquisition (Mondria & Wiersma, 2004; Webb, 2005, 2009). It was found that both types of instruction were effective in promoting word knowledge, but productive learning was more beneficial for learning the form than receptive learning whereas receptive learning led to higher gains in learning the meaning than productive learning. The findings were claimed to support the Transfer Appropriate Processing (TAP) theory (Morris, Bransford & Franks, 1977) which

states that learners' test achievement will be at their best when the retrieval processes in the tests match the learning processes in the tasks. However, the relative effects of input- or output-based instruction, compared to list learning, remain unclear. Therefore, an empirical study that explores the effects of input- and output-based tasks on vocabulary learning may be worthwhile. From a pedagogical perspective, this study is important because its findings might serve as useful references for language teachers to balance the tasks in a way that can optimize their effects in vocabulary teaching.

Research Questions

To this end, the following research questions were formulated:

- 1) Is there an effect of input-based and output-based tasks on vocabulary learning?
- 2) If so, is there a difference in the effectiveness of input- and output-based tasks on vocabulary learning?

Methodology

Participants

Sixty Vietnamese EFL university students (L1 = Vietnamese; aged 18 to 20 years) participated in this experiment. Participants were majoring in Tourism and Hospitality Management and were from two different universities in Vietnam. Participants were expected to be at the A2- B1 level defined by CEFR (Common European Framework of Reference). Participants from one university were assigned to the experimental group (n=30) and participants from the second university were assigned to the control group (n=30). To control for individual differences and differences between the experimental and control group, a vocabulary size test (developed by Nguyen & Nation, 2011) was used to get an estimate of learners' overall vocabulary knowledge in English. The analysis showed that there was no significant difference in vocabulary size between two groups ($t = .56, p = .25, df = 58, d = .14$).

Design

A pretest-posttest design was employed with treatment (experimental or control) as a between-participants variable and type of task (input- or output-based tasks) as a within-participants variable. To investigate the effect of the treatment (input- or output-based), the experimental group was asked to do the input-based as well as the output-based tasks, while the control group was not. The control group only took the tests. The control group was used to control for potential test effects (e.g., guessing based on the first letter of the target words, learning from pretests to posttests) or learning outside the treatment (e.g., looking up target items at home).

Tasks

We used a strategy called ‘task essentialness’, which means that the tasks were designed in such a way that the task outcome can only be satisfactorily achieved if the learners use the target items (Loschky & Bley-Vroman, 1993). Additionally, all tasks were designed as technology-supported tasks in which learners had to complete the tasks on the learning management system *Moodle*. All tasks were piloted to check whether the targeted items were indeed task-essential. If not, the tasks and prompts were improved.

Input-Based Tasks

Learners in the experimental group performed two input-based tasks. Learners were asked to (a) read L1 emails, (b) watch L2 captioned (i.e., with subtitles in the L2) videos about tourist attractions in famous places, and (c) write two reply e-mails in their L1.

In the first reply e-mail, participants had to give travel tips in Vietnamese to a friend, whereas in the second reply e-mail they had to suggest a tour schedule in Vietnamese to tourists. To ensure that participants would process the target words embedded in the L2 video, the participants were asked to read an L1 email before watching the videos. In this L1 email, meaningful questions related to the target items were asked. In the reply email, participants had to answer the questions from the L1 email using the information from the videos. Participants could not answer the questions/respond to the email if they did not comprehend the target items embedded in the videos. Learners did not have the opportunity to negotiate/ask for clarification with the teachers, but they could look up words in a web-based glossary (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 travelling company that 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 requested you to design a tour based on his requirements and write a reply email 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)...*

To design a tour schedule that takes into account the requirement 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 sunrise is one of the most incredible ways to see Cairns.”

Because the learners have to write a reply email in 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.

Figure 1

The web-based English-Vietnamese glossary (created with H5P Accordion tool)

| 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 |

There were in total 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, taken from *Viator Travel* Youtube channel which features videos, narrated by native-speakers, about well-known tourist destinations and attractions worldwide. The videos were piloted with a group of participants whose language profile was similar to those in this experiment to ensure that the video content was not too difficult. We used Nation’s Range software to analyze the lexical profile of the video content. The analysis revealed that the 4,000 most frequent word families provided a 95% coverage. While watching the videos, participants had access to a web-based English-Vietnamese glossary, which contained words taken from the videos (both target items and others) that might be unknown to the learners. Participants were free to rewind the video clips while performing the tasks. However, time-on-task was controlled for, as learners were informed that they had two hours for completing both input-and output-based tasks and probably needed 30 minutes at most to complete each task. Even though we did not control for frequency of exposure in the number of viewings,

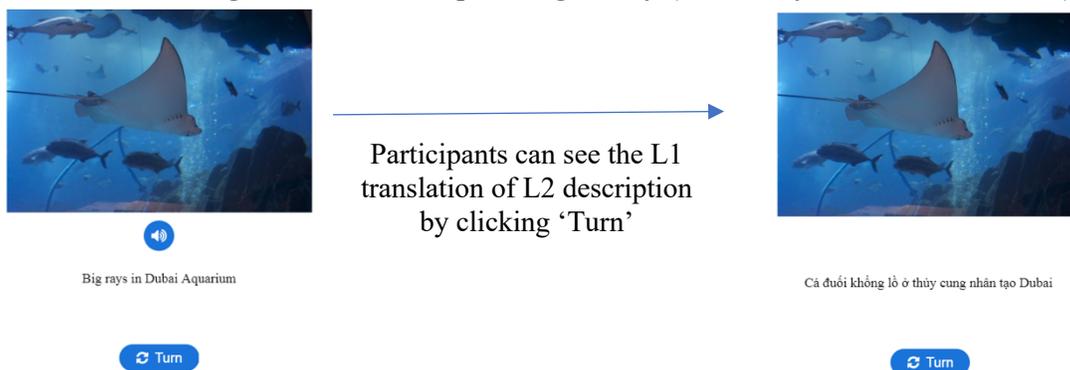
we assume that the frequency of exposure is roughly the same for most participants given the time constraints.

Output-Based Tasks

Participants in the experimental group also completed two output-based tasks: 1) write a travel blog in English based on Instagram photos, and 2) write an email to foreign tourists in English to propose a travel itinerary based on Vietnamese tourist leaflets in which information was presented in bulleted lists. Unlike in the input-based tasks, we did not provide English-language input through videos. We used picture prompts (i.e. Instagram photos) and L1 bulleted lists (in the leaflets) to elicit the use of target items. To facilitate learners' output production, we created a web-based English-Vietnamese picture glossary which is justifiable 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. The glossary contained pictures and L2 descriptions related to the target items in written and spoken mode (i.e. learners could see the descriptions and click on the loudspeaker symbol to hear the spoken form) (see Figure 2).

Figure 2

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



Target items

Twenty lexical items (10 single nouns, 10 compounds) were selected as target items. Target items were chosen based on the following criteria:

1. They had to be relevant to the field of tourism and travelling to meet the learners' needs because the participants of this study majored in English and Tourism Hospitality.
2. They should have a high level of concreteness (4-5) because pictures were used as prompts. Concreteness was checked by using Brysbaert et al.'s (2014) concreteness ratings. The concreteness ratings of the compounds were checked on the concreteness ratings of the head noun of the compounds viz. the word that determines the core meaning of the sequence (e.g., the head noun of *hot-air balloon* is *balloon*). We used COCA (Corpus of Contemporary American English) (Davies,

2009) to check the compounds' MI scores. MI score higher than three was aimed for since it is a commonly accepted cut-off score for an item to be considered a multi-word unit (McEnery, 2006).

3. The target items were cross-checked with Vietnamese teachers to guarantee that the items were not in the textbook or taught in class.
4. Given that most task-based research has focused on single words, we included single words as well as compounds (= a type of formulaic sequences) in our target items. Research has shown that formulaic sequences can be learned in deliberate learning activities (e.g., Peters, 2016; Webb & Kagimoto, 2009) and incidentally through reading (e.g., Pellicer-Sánchez, 2017; Webb et al., 2013). Yet, to our knowledge there has been little research on the learning of formulaic sequences (= compounds in our study) through tasks. Learners' knowledge of collocations tends to lag behind (Nguyen & Webb, 2017) also in Vietnamese contexts.

The frequency of encounters with the target items in the video clips was controlled: all items appeared only once (for a list, see Table 1). To avoid any confounding effect from the target items, the items were counterbalanced across the input- and output-based tasks. This means that half of the participants (15 students) processed half of the target items in the input-based tasks and the other half in the output-based tasks. The other 15 participants processed the second half of the target items in the input-based tasks and the first set of 15 items in the 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 |
| Armour | 4720 | 4.76 | Northern light | 14.00 | 4.21 |
| Sword | 7911 | 4.93 | Hot spring | 12.97 | 3.89 |
| Mosque | 4229 | 4.54 | Zipline | 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

We used Nguyen and Nation's (2011) bilingual version of the Vocabulary Size Test (VST) to control for individual differences in overall vocabulary knowledge in English. The VST is a frequency-based meaning recognition test which samples 10 words from 14 frequency bands of 1000 words (1K-14K). The test contains 140-items. Each item is presented in English and is accompanied by four options in Vietnamese (one correct

definition and three distractors). A pilot showed that the test procedure was too long, which is why we decided to develop a short version (70-item) of the bilingual test in the same way as Beglar (2010) did for the monolingual version: we randomly selected five items per frequency level from the 140-item test. The 70-item version of the bilingual test had good internal consistency (Cronbach's $\alpha = .85$, $n=60$).

Vocabulary Knowledge Tests

Learning of the target items was measured by means of four tests: a test focusing on spontaneous use of the target items, a form recall test, a meaning recall test, and a meaning recognition test. All tests were administered online in Moodle. The importance of using more than one test in vocabulary assessment has been widely emphasized in vocabulary research (Nation & Webb, 2001; Webb, 2009) because several word knowledge aspects can be tested. The four tests were administered in descending difficulty to avoid a testing effect: first the spontaneous use, then the form recall, third the meaning recall and finally the meaning recognition test. The same test battery was used for both pre- and post-test but the items in the tests were ordered differently.

- a. *Spontaneous use test*: In this test, participants were asked to write a short paragraph (maximum 250 words) to describe their travel experiences based on 20 pictures, corresponding to the 20 target items. The pictures were similar but not identical to the ones used in the treatment.

Example:

Watch the video and write an email to your friend (max. 250 words) to describe what you experienced/saw/visited in the five countries shown in the video. You have 30 minutes to finish the task. You are not allowed to use a dictionary or any other reference sources.

(You can start the email as follows)

Dear ...

How are you? Last summer, I visited/travelled towith my families/friends/ etc. In Japan, I saw/ experienced/ visited

- b. *Form recall test*: In this test, participants had to provide the word that best described a given picture, similar to the one given in the treatment. The first letter of the item was given as a clue. The test consisted of 25 items, 20 target items, and 5 distractors.

Example:

We skied together towards the c _ _ _ _ l _ _ _ (2 words) that would take us to the mountain.

- c. *The Meaning recall test* was an English-Vietnamese translation test. The items were presented in written as well as aural forms because we used captioned audio-visual input in the input-based tasks. The participants were asked to give the Vietnamese meaning or explanation of the English word/compound without context. The test consisted of 25 items, 20 target items, and 5 distractors.

Example:

Glass bottom boat _____

- d. *Meaning recognition test:* The participants had to select the meaning of the underlined target words. Target words were presented in low contextualized sentences to avoid meaning inference and test learning. The test included an “I don’t know” option to minimize guessing. The test contained 25 questions, 20 target items, and 5 distractors.

Example:Don’t let her touch the sword.

- a. thanh gươm (= sword)
- b. tài liệu (= document)
- c. chuông báo cháy (= fire alarm)
- d. bếp lò (= oven)
- e. I don’t know

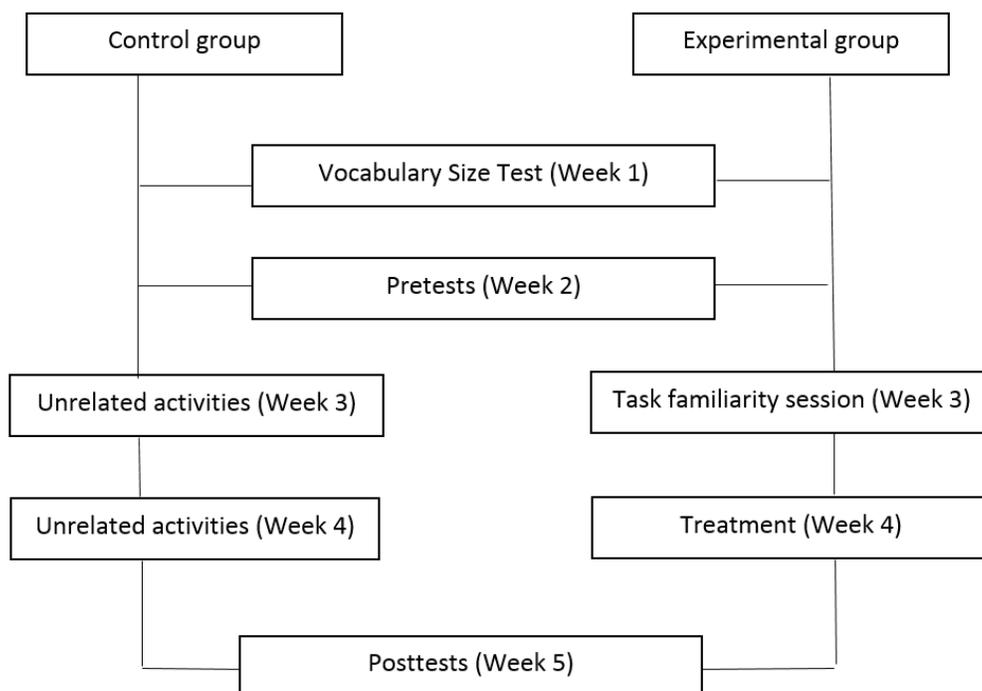
Procedure

The procedure consisted of five sessions. Given the flexibility in time and place for the learners, all data were collected via computer-based online tasks and tests on the learning management system *Moodle*. All learners participated on a voluntary basis. They did the tasks and tests on an individual basis at home. To ensure that the participants did the tasks and tests themselves without help from other people or other reference sources, they were asked to activate screen-and-webcam recording during the tests; and screen-recording while carrying out the tasks, using Flashback Express software – a tool which allows users to record activities on screen as well as their webcam. The participants had to send the recording files to the researcher immediately after they finished the tasks and the tests. The participants were trained to use the screen recording software and did a trial recording test two weeks before the experiment. Therefore, they were assumed to master the use of screen recording software in the treatment. They had two hours in total to finish the learning tasks, and seventy-five minutes to do the pretests/posttests. The participants were requested to finish the learning tasks and tests in the assigned week, but they were free to do the tasks and the tests anytime within that week

Both the experimental and control group took the Vocabulary Size Test in the first week. One week later in session 2, participants in the experimental group worked on a task familiarity session. In the third week, both groups (experimental and control) took the pretests in the following order: first the spontaneous use, then the form recall, third the meaning recall and finally the meaning recognition test. To control for a test effect, learners had to do an unrelated test between the form recall and the meaning recall test, viz. 10-question listening comprehension test. In addition, the test items in the different vocabulary posttests appeared one by one, and the participants could not return to the previous question pages to review their answers. In the fourth week, participants in the experimental group did the input- and output-based tasks. In the meantime, the control group worked on unrelated lessons which focused on grammar and reading. The control group was not exposed to the target items. One week later, the participants of both groups took the posttests. As in the

pretest, learners had to do an unrelated test between the form recall and the meaning recall test, viz. 10-question listening comprehension test.

Figure 3
Research Procedure



Scoring and Analyses

The maximum score was 20 for each test. The tests were scored dichotomously. Two different scoring systems were used for the spontaneous use and the form recall test: a lenient and a strict scoring system. The lenient scoring system allowed for the measurement of partial vocabulary knowledge. On this system, learners got 0 for an incorrect response and 1 for a fully correct response or a response with one letter omitted/one letter added/one wrong letter. For instance, such responses as ‘peecok’, ‘peacok’, ‘peacock’ were awarded 0, 1 and 1 respectively; or both ‘dophin’ and ‘dolphine’ were awarded 1. For compounds, such responses as ‘air balloon’, ‘hot-air ballon’, ‘hot-air balloon’ were awarded 0, 1, and 1 respectively. The strict scoring scale only awarded points to fully correct responses. Therefore, in the examples above, only ‘peacock’ and ‘hot-air balloon’ were given one point. In the meaning recall test, the participants obtained one point for each item of which they gave correct Vietnamese meaning or explanation. For questions about the meaning of compounds, the participants obtained one point only if they could provide a ‘whole-word meaning’ of the compound (Libben, 2006), not a word-by-word translation. In the meaning recognition test, the participants obtained one point for each correct selection.

The results were analyzed with SPSS 24. Because the assumptions of normality were violated, we used non-parametric tests to conduct the analyses. To gauge the effects of the

treatment (input- and output-based tasks) (Research Question 1), Mann-Whitney U tests were used to explore the differences in gains before and after the treatments. To examine differences in the effects between these task types on different vocabulary aspects (Research Question 2), we used Generalized Estimating Equation (GEE) analyses in SPSS to fit a repeated measured logistic regression. A GEE analyzes results at item level, that is, one observation per response, per participant, which was considered appropriate given the dichotomous scoring of the tests. All main parameters were entered into the model. Specifically, within-participants variables, ‘task type’ (input vs. output) and ‘item type’ (single vs. compound) were entered as the main factors and learners’ general vocabulary size as a covariate. Because we wanted to control for learners’ responses in the online tasks, we also used learners’ – ‘correct use in task’ as another main factor (i.e., whether the target item was used correctly or not in the tasks). We did not control for the learners’ use of the web-based glossary. However, a qualitative analysis¹ showed that when a target item was not known in the pretests and was not looked up, the item was not used in the tasks. Target items that were looked up, on the other hand, were also used in the tasks, except for one participant who looked up all items in the output-based task but used only one item in this task. Finally, non-significant variables were then removed from the model.

Results

Prior Vocabulary Knowledge

A t-test was used to compare the mean scores because data was normally distributed. The t-test revealed that there was no significant difference between the experimental and control group in terms of prior vocabulary knowledge (see Table 2), with $t = .56$, $p = .25$, $df = 58$, $d = .14$.

Table 2

Mean score and estimate in word family per group

| | Min | Max | Mean score (Max = 70) (SD) |
|---------------------|-----|-----|----------------------------|
| Experimental (N=30) | 7 | 42 | 22.93 (7.88) |
| Control (N=30) | 9 | 35 | 23.96 (6.00) |

Learners’ Responses to Target Items When Performing Input- and Output-Based Tasks

We verified learners’ task performance in the input-based and output-based tasks. In other words, we checked whether learners could comprehend and produce the target items in the tasks. A response was considered accurate if the learners wrote down the Vietnamese meaning of target items in the input-based tasks or if they wrote the correct English form of target items in the output-based tasks. Mean scores of the correct responses on each task type were compared (see Table 3) using a Wilcoxon Sign-Rank Test because the normal distribution was violated. The analysis showed no significant difference in learners’

¹ In the qualitative analysis, we checked the learners’ look-up behaviors manually.

response to target items in two task types ($p = .092$), which implies that the focus on form techniques used in two task types had similar effects. The descriptives in Table 3 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 3

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

| | Minimum | Maximum | Mean (Max=10) | SD |
|--------------------|---------|---------|------------------|------|
| Input-based tasks | 7 | 10 | 9.47 | .776 |
| Output-based tasks | 7 | 10 | 9.77 | .626 |

Research Question 1: Is There an Effect of Input-Based or Output-Based Tasks on Vocabulary Learning?

Given the limited number of target items ($n = 20$), we used learners' absolute gains and not relative gains to investigate whether the experimental and control group differed significantly. The absolute gains were calculated at the item level (= posttest score – pretest score per item) rather than at the level of the total score. We used the strict scoring system in the spontaneous use and the form recall test (see Scoring and analyses). The lenient scoring system was used later to verify whether the scoring could have affected the results, but that was not the case. The descriptive statistics for the vocabulary gains in four tests are shown in Table 4. Both groups performed better on the posttest than on the pretest. On average, the learners gained knowledge of 6 out of 18.7 items that could potentially be learned at the level of spontaneous use, 7 out of 16.4 items at the form recall level, and 8 out of 12.8 items at the meaning recall level.

The Mann-Whitney U test showed that there was a significant difference between the experimental and control groups in the three tests. The experimental group gained more vocabulary knowledge than the control group in the spontaneous use test ($Z = -6.33$, $p = .000$, $d = 2.54$), the form recall test ($Z = -6.68$, $p = .000$, $d = 3.75$), and the meaning recall test ($Z = -6.41$, $p = .000$, $d = 3.44$). Effect size values were interpreted as small = .2, medium = .5, and large = .8 to infinity (Cohen's 1988). The comparison between the two groups showed that the effect size was large, which means the treatment indeed had a significant effect on vocabulary learning. However, in the meaning recognition test, no difference was found between the two groups ($Z = -1.93$, $p = .053$). This was likely due to a test effect from taking multiple vocabulary tests before the meaning recognition test. Therefore, results in the meaning recognition tests were considered invalid and not analyzed further.

Table 4
Vocabulary gains in the experimental and control group

| | Spontaneous use | | | Form recall | | | Meaning recall | | | Meaning recognition | | |
|--------------------------|-----------------|----------------|----------------|----------------|-----------------|----------------|-----------------|-----------------|----------------|---------------------|-----------------|----------------|
| | Pretest | Posttest | Absolute gain | Pretest | Posttest | Absolute gain | Pretest | Posttest | Absolute gain | Pretest | Posttest | Absolute gain |
| | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| Experimental (N = 30) | 1.30 (1.26) | 7.40 (3.30) | 6.10 (2.90) | 3.13 (2.30) | 11.90 (3.89) | 8.33 (2.82) | 7.23 (3.46) | 15.37 (3.17) | 7.93 (2.85) | 16.43 (3.02) | 18.93 (0.90) | 2.66 (2.73) |
| Control (N = 30) | 0.27 (0.44) | 0.70 (0.74) | 0.56 (1.04) | 3.17 (1.57) | 3.73 (1.61) | 0.56 (0.77) | 11.23 (3.38) | 12.23 (3.56) | 1.03 (1.67) | 16.17 (3.27) | 17.43 (3.03) | 1.26 (1.43) |

Table 5
Vocabulary gains in the experimental group

| | Spontaneous use | | | Form recall | | | Meaning recall | | | Meaning recognition | | |
|--------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|---------------------|----------------|----------------|
| | Pretest | Posttest | Absolute gain | Pretest | Posttest | Absolute gain | Pretest | Posttest | Absolute gain | Pretest | Posttest | Absolute gain |
| | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| Input-based | 0.63 (0.76) | 3.60 (2.04) | 3.00 (1.81) | 1.73 (1.28) | 5.47 (2.27) | 3.50 (1.88) | 3.53 (1.87) | 8.03 (1.56) | 4.50 (1.75) | 8.27 (1.38) | 9.63 (0.55) | 1.47 (1.43) |
| Output-based | 0.66 (0.80) | 3.80 (1.95) | 3.13 (1.79) | 1.40 (1.38) | 6.43 (2.26) | 4.83 (1.85) | 3.73 (2.62) | 7.33 (2.26) | 3.60 (1.71) | 8.17 (2.03) | 9.30 (0.70) | 1.37 (1.77) |

Table 6
Number and percentage of correct and incorrect responses in the posttests

| | | Incorrect responses | Correct responses | Total |
|-----------------|------------------|---------------------|-------------------|--------------|
| Spontaneous use | All target items | 377 67.1 % | 185 32.9% | 562 100 % |
| | Single words | 180 67.9% | 85 32.1% | 265 100% |
| | Compounds | 197 66.3% | 100 33.7% | 297 100% |
| Form recall | All target items | 242 48.7 % | 255 51.3% | 497 100 % |
| | Single words | 119 56.1% | 93 43.9% | 212 100% |
| | Compounds | 123 43.2% | 162 56.8% | 285 100% |
| Meaning recall | All target items | 139 36.9% | 238 63.1% | 377 100% |
| | Single words | 68 38.6% | 108 61.4% | 176 100% |
| | Compounds | 71 35.3% | 130 64.7% | 201 100% |

Research Question 2: Is There a Difference in the Effectiveness of Input- and Output-Based Tasks on Vocabulary Learning?

To answer this question, the GEE analysis was conducted on the posttest scores of items that were unknown to the learners in the pretests, viz. the items that could potentially be learned. Table 5 showed the descriptive statistics for vocabulary gains in input- and output-based tasks. It can be seen that there were gains from pretests to posttests in the input-based as well as in the output-based tasks.

Spontaneous Use Test

As can be seen in Table 5, three items were learned on average in both the input-based and output-based tasks. The GEE analysis, which was computed for 562 observations (see Table 6) revealed only one significant predictor, viz. learners' correct use of the item in the tasks themselves (see Table 7). Learners were 11 times more likely to use a target item in the spontaneous use test if they had used it correctly in either task ($1/.084 = 11.90$). Second, the analysis showed that task type, lexical type (single word or compound) and prior vocabulary knowledge were not significant predictors of the posttest scores in

spontaneous use test. The effect was the same with the lenient scoring system: only one significant predictor, viz. learners' correct use of the items in the tasks ($p = 0.10$)

Table 7

Effect size estimates from GEE for the spontaneous use posttest

| Parameter | Wald chi-square | df | Sig | B | Exp(B) | CI | |
|------------------------------|-----------------|----|-------|--------|--------|-------|-------|
| | | | | | | Lower | Upper |
| Intercept | 27.52 | 1 | <.001 | -.654 | .520 | .407 | .664 |
| Incorrect = 0 Correct = 1 | 5.952 | 1 | .015 | -2.481 | .084 | .011 | .614 |

Note. Exp(B) = odds ratio.

Form Recall Test

The analysis was run for 497 cases (see Tables 6 and 8). The analysis showed that words learned in the output-based tasks were recalled slightly better than words in input-based tasks, 4.83 versus 3.50 words on average (see Table 5). The analysis also revealed that there was a significant difference between input-based and output-based tasks on learners' form recall posttest scores (see Table 8). Learners were 68% ($\text{exp}(B) = 1.681$) more likely to produce target items correctly in the form recall test if the items were offered in the output-based tasks. The factors 'correct use in tasks', and prior vocabulary knowledge did not contribute significantly to the model. Further, compounds were learned better than single words in the form recall test. Learners were 70% ($\text{exp}(B) = 1.705$) more likely to produce the items correctly in the form recall test if the items were compounds. No interaction effect between the task type and the item type was found. The effects were the same for the lenient scoring system: posttest scores of input and output tasks were significantly different ($p = .007$); items offered in the output tasks were produced more correctly in the test ($\text{exp}(B) = 1.684$); and compounds were recalled 2 times better than single words in the test ($\text{exp}(B) = 2.165$).

Table 8

Effect size estimates from GEE for the form recall posttest

| Parameter | Wald chi-square | df | Sig | B | Exp(B) | CI | |
|----------------------------|-----------------|----|------|-------|--------|-------|-------|
| | | | | | | Lower | Upper |
| Intercept | 5.663 | 1 | .000 | -.518 | .596 | .389 | .913 |
| Output = 0 Input = 1 | 7.541 | 1 | .006 | .520 | 1.681 | 1.160 | 2.436 |
| Compound = 0 Single = 1 | 11.451 | 1 | .001 | .534 | 1.705 | 1.252 | 2.322 |

Note. Exp(B) = odds ratio.

Meaning Recall Test

There were higher gains for input- than for output-based tasks in the meaning recall test (4.50 vs. 3.60 words) (see Table 5). The analysis, which was run for 377 cases (see Table 6 and 9), showed that there were two significant predictors: task type and correct use

in the tasks. Learners were 1.7 times more likely to recall a target item's meaning correctly if it was offered in input-based tasks ($1/.589 = 1.7$). Also, learners were almost three times more likely to provide a correct response if the word had been used correctly in the tasks. ($1/.337 = 2.96$). Lexical type and prior vocabulary knowledge did not make a significant contribution to the model.

Table 9

Effect size estimates from GEE for the meaning recall posttest

| Parameter | Wald chi-square | df | Sig | B | Exp(B) | CI | |
|------------------------------|-----------------|----|------|------------|--------|-------|-------|
| | | | | | | Lower | Upper |
| Intercept | 21.027 | 1 | .000 | .875 | 2.400 | 1.651 | 3.489 |
| Output = 0 Input = 1 | 8.715 | 1 | .003 | -.529 | .589 | .415 | .837 |
| Incorrect = 0 Correct = 1 | 7.175 | 1 | .007 | - 1.087 | .337 | .152 | .747 |

Note. Exp(B) = odds ratio.

Discussion

Input- and Output-Based Tasks and Their Effects on L2 Vocabulary Acquisition

In answer to the first research question, results indicated that both input- and output-based tasks were effective in enabling the learners to acquire word knowledge. Our findings show that there was a considerable amount of learning from both input- and output-based tasks, especially since learning was measured one week after the learning sessions. The findings also showed that input-based tasks resulted in gains in meaning recall as well as form recall and so did the output-based tasks. It is not surprising to find that the output-based tasks can enhance the meaning recall as well as the form recall aspect because the task design has provided the learners with substantial opportunities to produce the L2 forms and process an amount of input, albeit to a limited extent. The effects of output-based tasks might also be interpreted in the light of how the task design triggered the learners' cognitive processes. In this study, the output tasks were designed to push the learners to engage in cognitive processing of prompted pictures/L1 texts through which they were alerted of their knowledge gaps which could then be filled by attending closely to input in the web-based glossary.

As for the input-based tasks, our findings were in line with previous studies (Erlam & Ellis, 2018; Shintani, 2012) which reported that input-based tasks could successfully facilitate vocabulary acquisition receptively as well as productively. The learners' responses in input-based task performance (Table 3) demonstrated that the participants successfully processed the L1 meaning of target items, which explains their gains in the meaning recall test. Yet, learners' gains in the form-recall test are puzzling. Even though L2 production was not prohibited in the input-based tasks, video recordings showed that the participants did not engage in written L2 forms production or any form of interaction with the teacher. Therefore, it seems unconvincing to explain the learners' gains in form recall as the result of interactional negotiations as in Shintani (2011, 2013)'s studies. In the present study, the qualitative analysis of video recordings showed that although being

given the same amount of time for each task (30 minutes), some participants needed less time to finish the input-based tasks than the output-based tasks. These participants tended to use the little extra time left to revise the writing task or rewind the videos or look up the web-based glossary, which might have provided them with opportunities to process the input more deeply.

Differential Effects Between Input- and Output-Based Tasks

In answer to research question 2, which asked whether there was a significant difference between the effects of input- and output-based tasks on vocabulary learning, the answer is affirmative. Although the input-based tasks led to learning in all tests (learning the meaning as well as the form), they resulted in larger gains in meaning recall. Similarly, the output-based tasks led to learning in all lexical aspects, but the larger gains were found in the form recall test. These findings were not in line with earlier findings in Shintani's studies (2011), which found that PPP activities and input-based tasks had similar effects on learning nouns (Shintani, 2011).

A plausible explanation for the results might be based on the Transfer Appropriate Processing (TAP) theory (Morris et al., 1977) which holds that learners' test achievement will be at their best when the learning processes match the retrieval processes in the tests. In our study, the learners' higher scores for input-based tasks items in the meaning recall test could be explained by the fact that the learners attended primarily to the meaning of target items in the input tasks. Similarly, the learners' primary attention to the form of target items in the output tasks could facilitate their form recall of these items. The findings are also consistent with previous research on learning word pairs (Mondria & Wiersma, 2004; Webb, 2009), which found that receptive learning of word pairs was better suited for developing receptive knowledge (=meaning recall) while productive learning of word pairs led to greater gains in productive word knowledge (=form recall).

In addition to the task effects, the analysis showed that the nominal compounds were learned significantly better than the single words at the form recall level. A possible reason is that the compounds consist of relatively high-frequency constituent words, which might have made them easier to recall. Also, given that the concreteness level of target items was controlled for, we assume that it could be likely due to the perceived novelty of compounds in terms of either semantics or morphology compared to the single words that make them relatively salient for the learners. Since compounds are a special type of multiword units whose meaning is considerably concrete, further research on other types (e.g., collocations, phrasal verbs) is needed in order to refine our understanding of whether and to what extent lexical types impact the task effects.

It should be noted that there are a number of limitations in this study. The first one concerns the small number of target items as well as the short period of instruction, which impacts the generalizability of the findings. Another limitation is that we did not control the learners' use of strategies when watching the video (e.g., pausing, rewinding or freeze-framing), which might influence the frequency of exposure to target items. Future studies might explore how these strategies moderate the effects of captioned videos on language acquisition. Also, this study exclusively focuses on writing tasks and written vocabulary knowledge. Given that writing and speaking are assumed to pose different demands on cognitive involvement which might lead to differences in language acquisition (Grabowski,

2007; Halliday, 1989), further research focusing on speaking tasks and spoken vocabulary knowledge is needed. In addition to learning outcomes, future research should examine learners' cognitive processes during the task performance as well, preferably using intrusive techniques (e.g., think-aloud protocol, eye-tracking, keystroke logging) and/or non-intrusive ones (e.g., stimulated recall).

Conclusion

This study has investigated the comparative effects of input- and output-based tasks on vocabulary acquisition by university EFL learners. The study confirms that when tasks (whether input-based or output-based) can create a functional need for learners to attend to target items, vocabulary acquisition can take place. Second, both task types were found to successfully promote receptive and productive word knowledge, but input-based tasks led to more gains in meaning recall and output-based tasks showed higher gains on the form recall test. These findings suggest that task-based instruction, realized through focused input- and output-based tasks may be well suited to provide opportunities to focus on form in L2 classroom and improve L2 learning, particularly for learning vocabulary.

In terms of pedagogical implications, together with previous studies (e.g. Erlam & Ellis, 2018; Shintani, 2012), this study has shown that input-based tasks can be successfully implemented to foster vocabulary learning. Thirty university students of English demonstrated successful acquisition of both single words and compounds after doing the tasks. This study also suggests that teachers can employ input-based tasks to teach vocabulary to A2-B1 level students. Also, input-based tasks can be operationalized in a different way other than listen-and-do tasks to teach vocabulary. The study also has implications for the design of output-based tasks. Picture-promoted and L1 text-promoted writing tasks employed in this study could be useful in vocabulary or writing class as an additional activity that can control the learners' focus of predetermined target language without ruling out the meaning. Further, the comparative findings indicate that both task types are beneficial but the input-based tasks seem to be better for learning the meaning of new lexical items whereas the output-based tasks are more effective to develop productive vocabulary knowledge. These findings suggest that each task type may affect different aspects of vocabulary knowledge. Based on these findings, teachers can balance vocabulary activities to ensure the development of well-rounded, usable vocabulary knowledge. It might thus be useful to administer input-based tasks at first to provide learners with opportunities to comprehend the meaning of target items in context before fostering productive vocabulary knowledge by means of output-based tasks.

Correspondence should be addressed to Phuong Thao Duong.
Email: duongthao204@gmail.com

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

- Brybaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911.
- 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>
- De la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary. The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, *24*, 81-112. doi:10.1017/S0272263102001043
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994) Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, *44*, 449-491.
- Ellis, R. & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, *21*, 285-301. doi:10.1017/S0272263199002077
- Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus-on-form. *System*, *30*(4), 419-432. [https://doi.org/10.1016/S0346-251X\(02\)00047-7](https://doi.org/10.1016/S0346-251X(02)00047-7)
- Ellis, R. (2009). Task-based language teaching: sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- 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. doi:10.1017/S0261444818000216
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., ... Yoon, H.J. (2018). Incidental vocabulary learning in a natural reading context. An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563-584.
- Grabowski, J. (2007). The Writing Superiority Effect in the Verbal Recall of Knowledge: Sources and Determinants. In Rijlaarsdam, G. (Series Ed.) and M. Torrance, L. van Waes & D. Galbraith (Volume Eds.), *Writing and Cognition: Research and Applications*, Vol.20, pp.165-179. Amsterdam: Elsevier
- Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford University Press.
- Izumi, S. (2002). Output, input enhancement, and the Noticing Hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, *24*, 541-577.
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, *24*, 168-196.
- Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly*, *34*, 239-278.
- Keck, C., Iberri-Shea, G., Tracy-Ventura, N., Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91-131). Philadelphia: John Benjamins.
- Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *Modern Language Journal*, *92*(1), 114-130.
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of

- comprehensible input. *Applied Linguistics*, 4, 126-141.
- 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). Clevedon: Multilingual Matters.
- Laufer, B. (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook* 5(1), 223-250. Retrieved from <https://muse-jhu-edu.kuleuven.ezproxy.kuleuven.be/article/204198>
- Libben, G. (2006). Why study compound processing? An overview of the issues. In Libben, G., & J. Gonia (Eds.), *The Representation and Processing of Compound words* (pp.18-39). Oxford University Press.
- McEnery, T. (2006). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Abington, UK: Routledge.
- 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
- Mondria, J. & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 79-100). Benjamins.
- 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](https://doi.org/10.1016/S0022-5371(77)80016-9)
- Nation, I. S. P. & Webb, S. A. (2011). *Researching and Analyzing Vocabulary*. Heinle, Cengage Learning.
- Nation, I. S. P., & Heatley, A. (1994). Range: A program for the analysis of vocabulary in texts [software]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>
- Nassaji, H. & Tian, J. (2015). 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, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal*, 42(1), 86-99. <https://doi-org.kuleuven.ezproxy.kuleuven.be/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.
- Nguyen, T.M.H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298-320.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, 38(1), 97-130.
- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381-402. <https://doi.org/10.1177/1362168815618428>
- Peters, E. (2016). The learning burden of collocations: the role of interlexical and intralexical factors. *Language Teaching Research*, 20(1), 130-138. <https://doi.org/10.1177/1362168814568131>
- 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. doi: 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.
- Révész, A. (2007). Focus on form in task-based language teaching: Recasts, task complexity, and L2 learning (Doctoral dissertation). Retrieved from Dissertations & Theses: A & I.
- Révész, A. (2017). Replication in task-based language teaching research: Kim (2012) and Shintani (2012). *Language Teaching*, 1-11. Doi:10.1017/S0261444817000283
- Révész, A. & Han, Z-H. (2006). Task content familiarity, task type and efficacy of recasts. *Language Awareness*, 15, 160-179. doi: 10.2167/la401.0
- Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-22). New York: Cambridge University Press.
- 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. doi: 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. doi: 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. doi: 10.1002/tesq.54
- 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, Cambridge.
- Swain, M. (2005). The output hypothesis: theory and research. In Hinkel, E. (Ed.), *Handbook on Research in Second Language Teaching and Learning*. Lawrence Erlbaum.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16, 371-391.
- Van den Branden, C. (1997). Effects of negotiation on language learners' output. *Language Learning*, 47(4), 589-636.
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27, 33-52. doi:10.1017/S0272263105050023
- Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, 40(3), 360-376. <https://doi.org/10.1177/0033688209343854>
- Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43(1), 55-77. <https://doi.org/10.1002/j.1545-7249.2009.tb00227.x>
- Webb, S., Newton, J., & Chang, A. (2013). Incidental Learning of Collocation. *Language Learning*, 63(1), 91-120. doi: 10.1111/j.1467-9922.2012.00729.x

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.