Learning Formulaic Sequences at a Japanese University: Flashcards, Tasks, and Speaking Practice

２０２３年９月

北九州市立大学大学院社会システム研究科

博士（学術）学位請求論文

椎葉　彩

論文要旨

英語の句動詞やコロケーションといった定型表現は、自然な英語の使用において重要な役割を果たし、複雑な考えを正確かつ流暢に表現することを可能にしている。しかし、定型表現の習得は、意味が文字通りでないなどの理由から困難を伴うことがある。それにも関わらず、英語学習者が句動詞やコロケーションを効果的に習得する方法については、まだ多くの未解明な点が存在しており、教育者や学習者にとって有益な指針が必要であると考えた。したがって、本論文では、日本の大学生の句動詞とコロケーションの習得について研究を行った。フラッシュカード学習による検索練習を中心とした学習アプローチの有効性を探るとともに、スピーキング学習が学習成果に与える影響についても検討した。その結果、以下の注目すべき新たな知見が得られた。

1．句動詞のフラッシュカード学習を行った結果、句動詞を長期間学習しなかった後でも、顕著な記憶保持率を達成した。
2．学習タスクの種類や参加頻度は、句動詞の学習成果にほとんど影響を及ぼさなかったが、タスクに費やした総時間はわずかに影響を示した。
3．句動詞を使用したスピーキング学習を取り入れることにより、自然発話における句動詞の使用率が増加した。
4．コロケーションを使用したスピーキング学習により、スピーキングの流暢さが向上した。

これらの結果は、定型表現の習得において効果的な学習戦略を模索する教育者や学習者にとって重要な指針を示すものである。

ABSTRACT

English formulaic sequences (FSs), including phrasal verbs (PVs) and collocations, hold a crucial position in natural language usage, enabling both native and non-native speakers to express complex ideas fluently and accurately. Although mastering FSs poses challenges for learners due to several factors, such as their non-literal meanings, many aspects of how English learners can effectively acquire PVs and collocations remain unexplored. Thus, this thesis aimed to investigate the acquisition and retention of PVs and collocations among Japanese university students. It explored the effectiveness of various learning approaches centered on computer flashcard programs and retrieval practice, while also examining the impact of speaking tasks on learning outcomes and speaking fluency. The findings revealed noteworthy insights: (1) learners achieved remarkable memorization rates for PVs, even after an extended disuse period; (2) although task type and frequency of participation exhibited minimal influence, the total time invested in a variety of learning tasks demonstrated a modest impact; (3) incorporating speaking activities using PVs proved beneficial, resulting in increased usage of PVs in spontaneous speech; (4) speaking activities involving collocations showcased improvements in learners' speaking fluency. These findings carry important implications for language educators and learners seeking effective strategies in vocabulary acquisition.

TABLE OF CONTENTS

CHAPTER 1. Introduction

This chapter provides an introductory overview of the background, purpose, rationale, and theoretical framework of the present study. Additionally, the chapter outlines the overall structure of the thesis.

1.1 Background

This thesis explores English formulaic sequences (FSs), examining the domain of phrasal verbs (PVs) and collocations as encountered by Japanese university students over the course of an academic semester. According to Erman and Warren (2000), FSs make up 55 percent of a text. The prevalence of PVs becomes even more significant, as learners encountering them approximately once every 150 words, as uncovered by the corpus analysis by Gardner and Davies (2007). The significance of PVs can be seen in both spoken and written discourse, as evidenced by the studies conducted by Alali and Schmitt (2012) and Le-Thi et al. (2017).

Looking at the area of vocabulary learning, learners may encounter the pivotal role of memorization. It is within this realm that the power of flashcards emerges, providing a popular method that allows learners to engage in repeated exposure to target vocabulary, strategically interleaved with well-timed delays, thereby propelling the process of memorization (e.g., Kornell, 2009; Nakata, 2015, 2017). Thus, a notable gap

arises between the limited exploration of the effectiveness of spaced retrieval practice, and a widely used memorization technique, in the context of FS learning.

1.2 Purpose and Rationale

The purpose of the thesis is to investigate the acquisition and usage of FSs in language learning. As FSs constitute a significant part of language use, understanding the process of their acquisition and usage is crucial. However, it is well-known that vocabulary acquisition, including FSs, is a time-consuming endeavor.

The decision to adopt a longitudinal approach in this thesis is motivated by the gradual nature of language acquisition and the need for comprehensive studies in the field of second language acquisition (SLA). Scholars in SLA have long emphasized the importance of conducting in-depth studies that span an extended period, allowing for the diligent observation and tracking of students' progress over time (Ortega & Brynes, 2008; Ortega & Iberri-Shea, 2005). While acknowledging that the duration of an academic semester may not be extensive in the realm of memory testing, it is worth noting that this timeframe exceeds that of many other studies in the field (e.g., Karpicke & Blunt, 2011; Kornell, 2009; Nakata, 2008).

1.3 Theoretical Framework

Formulaic Sequences

Formulaic sequences (FSs) encompass several types of expressions, including idioms, phrasal verbs, lexical bundles, and collocations. They are recognized as crucial components for achieving native-like fluency in a language (Sinclair, 1991; Wray, 2000).

However, studies consistently indicate that learners of English as LX learners encounter challenges in producing FSs as proficiently as native speakers (Granger, 1998; Hyland, 2008; Li & Schmitt, 2009; Wray, 2000), and acquiring native-like "formulaicity" poses a complex challenge for them (Ellis, 2012).[1] Furthermore, LX learners often tend to avoid using FSs, potentially due to differences between their first language and English (Littlemore, 2003).

These difficulties and disparities in the use of FSs between native and non-native speakers emphasize the need for further investigation into the acquisition and use of FSs, specifically within the context of English as a second language. By exploring the

[1] Dewaele (2018, p.3) "suggested using the label 'LX,' meaning any foreign language acquired after the age at which the first language(s) was acquired, that is after the age of 3 years, to any level of proficiency. It is then possible to be either specific and compare the person's L2, L3, or L4, or to make a more global statement about the person's LXs." I would also like to refer to a language that learners study as LX.

acquisition and use of FSs this thesis aims to contribute to our understanding of the challenges faced by LX learners and their progress in developing FS competence.

Phrasal Verb List

One of the major decisions in this thesis was determining which PVs to include in the vocabulary learning program. While there is a significant difference between non-native and native speakers in terms of their knowledge of PVs (Garnier & Schmitt, 2016; Schmitt & Redwood, 2011), the selection of PVs for inclusion in a vocabulary learning remains an important consideration.

Various lists of PVs based on corpus analysis have been developed to provide guidance in selecting PVs for instructional purposes. The PVs included in this thesis were derived from the PHaVE List (Garnier & Schmitt, 2014), which is considered the most comprehensive and up-to-date list of PVs.

Retrieval Practice

The concepts of memory retention, retrieval practice, and spaced learning play a crucial role in the theoretical framework of this thesis. These concepts inform the design and methodology of the study, focusing on the effectiveness of spaced learning in the acquisition and retention of English FSs.

Ebbinghaus (1993) emphasized the importance of retrieval practice in long-term memory retention, demonstrating that memories gradually fade and are forgotten without active retrieval. In the context of vocabulary learning, the use of retrieval practice and spaced learning becomes essential for effectively learning and retaining a large number of words (Kornell, 2009; Nakata, 2013).

While prior research has explored the effectiveness of retrieval practice and spaced repetition in vocabulary learning, little is known about the application of spaced learning in the acquisition of English FSs over an extended period. The inclusion of spaced learning strategies in the study of FSs is expected to result in enhanced retention for learners, surpassing the effectiveness of traditional massed learning methods.

1.4 Thesis Organization

The thesis is organized into four main experiments. The organization of the thesis is as follows:

- Chapter 2. Experiment 1: Learning Phrasal Verbs in Vocabulary Learning Program

This chapter describes the establishment of a vocabulary program for learning PVs. The efficacy of flashcard learning with spaced retrieval practice is investigated, along with a delayed post-test conducted after one year or six months to assess the retention of PVs.

The results demonstrate the effectiveness of flashcard learning in acquiring and retaining PVs.

- Chapter 3. Experiment 2: Learning Phrasal Verbs with Tasks

This chapter explores the effectiveness of online tasks in addition to flashcard learning for improving learners' acquisition of PVs. The study examines the impact of different types and frequencies of the tasks on learning outcomes. The results reveal that, contrary to expectations, the type of tasks did not have a considerable influence on learners' achievement in acquiring PVs. However, it was observed that the total time devoted to these activities played a role in the learning process, albeit to a lesser extent.

- Chapter 4. Experiment 3: A Narrative Task Approach for Learning Phrasal Verbs

This chapter focuses on the learners' ability to spontaneously increase their usage of PVs. The study spans ten weeks and employs a story description approach, prompting learners to create a story each week that incorporates target PVs while describing pictures in a cartoon story. The results indicate that participants engaged in picture description activities demonstrate a greater spontaneous use of PVs compared to those who did not engage in such activities.

- Chapter 5. Experiment 4: Speaking Fluency with the Use of Collocations

This chapter investigates the improvement in learners' speaking fluency after studying collocations using flashcard learning and engaging in speaking activities. The results reveal that learners who actively participate in collocation learning and speaking activities experience significant improvements in their speaking fluency.

- Chapter 6. Concluding Remarks and Future Directions

This chapter provides a concise overview of the key findings from Experiments 1 to 4 and discusses the broader implications of the experiments. It also outlines the limitations inherent in the studies and suggests potential avenues for future research. In conclusion, this thesis makes a significant contribution to the field of SLA by providing empirical evidence on the effectiveness of diverse approaches to learning FSs, particularly PVs and collocations. The findings offer valuable insights for language teaching and learning practices concerning these vital language components.

1.5 Chronological Progression of Research

As the final section of the introduction, I provide an overview of the chronological progression of my research. The research began with Experiment 1, which marked the initial exploration during my Ph.D. studies. The objective of Experiment 1 was to

develop a vocabulary program that incorporated academic vocabulary and PVs at the university level. Drawing upon previous research on vocabulary acquisition, I recognized the effectiveness of spaced repetitions and subsequently implemented a flashcard system within Experiment 1.

Building upon the findings of Experiment 1, I proceeded to Experiment 2 with the aim of further enhancing PV learning by introducing additional tasks. This phase involved enthusiastic creation of short animations, referred to as "PVgif," serving as visual aids for all 120 PVs, in collaboration with my supervisor. The incorporation of visual elements was expected to enhance the enjoyment of PV learning and contribute to improved retention. However, despite the dedicated efforts, including the development of PVgifs, the additional tasks did not significantly influence learners' performance. On the other hand, it was observed that the total time spent on the tasks yielded a modest positive effect on learning outcomes.

Within the context of Experiment 2, it became apparent that the research had primarily focused on input-based PV learning, overlooking the importance of output and its contribution to learners' speaking fluency. This realization led to the exploration of Experiment 3, where I aimed to address this gap by introducing a speaking component into the learning process. According to prior empirical research, I acknowledged that formulaic sequences (FSs) played a crucial role in enhancing

learners' speaking fluency. To facilitate this, I created stories for learners to describe using target PVs and sought the assistance of an illustrator. The valuable experience gained from Experiment 3 set the stage for Experiment 4, where I expanded my research to investigate the integration of FSs into a speaking component. The objective was to improve learners' speaking fluency, with a specific focus on collocations. Consequently, I examined whether the combination of flashcards and speaking activities could lead to significant improvements.

CHAPTER 2.

Experiment 1: Learning Phrasal Verbs in Vocabulary Learning Program

Learning English phrasal verbs (PVs) can be a daunting task for language learners, particularly when it comes to selecting the appropriate particle. Have you ever struggled to make sense of PVs like *hold up* or *run into*? PVs are crucial for achieving a native-like level of English proficiency as they are an essential component of formulaic sequences. However, research has shown that English language learners often struggle with acquiring PVs (Garnier & Schmitt, 2016).

What makes PVs particularly tricky is that their meanings are not always obvious from their literal definitions. For instance, *hold up* means 'to prevent', which cannot be easily deduced from the component verb and particle. Similarly, *run into* and *come across* both mean 'to encounter by chance,' which cannot be understood from their literal meanings. As an English language learner, I found it impossible to guess the exact meaning from such meaning combinations of a verb and adverbial particle. It is therefore evident that learning PVs is critical for learners to comprehend more native-like English.

To tackle this issue, the PHaVE List (Garnier & Schmitt, 2014) provides a list of the 150 most frequent PVs with their key meanings. However, to date, no study has developed a practical PV learning program utilizing this list. Therefore, this thesis aims

to design a program for learning PVs using the PHaVE List and to investigate its effectiveness in enhancing PV acquisition among university students. The study will explore the optimal structure for a PV learning program and assess the number of PVs that can be learned over one academic semester.

2.1 Literature Review

What Phrasal Verbs Are

Phrasal verbs (PVs) are one type of formulaic sequences (FSs). FSs are crucially important elements in both spoken and written discourse (Gyllstad & Schmitt, 2019). It is important to understand their significance and complexities in language learning.

PVs are composed of a verb and an adverbial particle (e.g., *up*, *down*, *out*, *in*), which can drastically change the meaning of the verb (Gardner & Davies, 2007; Liu, 2011). For example, *put* has a different meaning when combined with *up* (*put up*) than it does with *out* (*put out*). Due to the complexities involved in learning PVs, many language learners struggle with their acquisition and usage.

They are also very important for learners because of being prevalent in everyday spoken language, informal contexts, and even in academic settings, although they are less common in writing-based registers like magazines, newspapers, and academic writing (Liu, 2011). Gardner and Davies (2007, p.347) note that learners may encounter one PV in every 150 English words on average. PVs are just one part of a large number of FSs that exist in English; it has been hypothesized that the number of FSs that exists in a language may exceed the number of single words in the lexicon (Jackendoff, 1995; Pawley & Syder, 1983). On the other hand, PVs are also considered to be a fundamental and essential component of verbal expressions equally to single words. Although they

are common in use, PVs are composed of a complex multiple-word structure (Gardner & Davies, 2007; Liu, 2011; Rudzka-Ostyn, 2003).

It is worth noting that many PVs have lexical equivalents or counterparts. For instance, 'postpone' can be considered a lexical equivalent for *put off*, 'execute,' 'conduct,' or 'perform' for *carry out*, and 'happen' or 'occur' for *come about*. However, PVs possess unique meanings that differentiate them from lexical verbs. 'Postpone' often implies the rescheduling of a formal or planned event, while *put off* tends to convey a sense of procrastination or avoidance. Due to these nuanced meanings and the varied usage patterns of PVs, learners often face challenges in comprehending and using them appropriately in context.

While the acquisition of FSs has been becoming more commonly studied for its strong contribution to speaking fluency (Boers et al., 2006; McGuire & Larson-Hall, 2017; Schmitt, 2010; Wood, 2007), few research studies have investigated the actual implementation of PVs within a vocabulary learning program.

Learning of Phrasal Verbs

One cognitive linguistic approach that has been gaining attention as an effective teaching and learning strategy for PVs is the use of conceptual metaphors. This approach is based on the work of Lakoff and Johnson (1980), who proposed a cognitive

approach to language by noting that some words extend their literal meanings to more metaphorical ideas.

A study by Yasuda (2010) used such conceptual metaphors to help students learn PVs. Yasuda presented 115 Japanese university students who were learning English with 21 PVs in one class period. The verbs included only those with the prepositions *up*, *down*, *into*, *out*, and *off*. For example, the PVs *dry up*, *open up*, *show up*, and *use up* were included. The control group (N = 56) saw the translations of the PVs and were told to memorize them. The experimental group (N = 59) were explicitly taught the orientational metaphors of the adverbial particles, such as "more visible/accessible is up," and shown how these metaphors helped understand the compositional meanings of the PVs. This group examined PVs grouped by adverbial particles and also looked at Japanese translations of the verbs. Both groups spent 10 minutes on the learning portion of the experiment. The students were then asked to fill in missing adverbial particles for 30 PVs embedded in sentences, half of which the students had seen (exposed PVs) and the other half which they had not (unexposed PVs) during the instruction. Results from this study showed that for the unexposed PVs (N = 15) the experimental group correctly filled in M = 1.25 more PVs than the control group, which was a medium effect size (Cohen's d = 0.68). Yasuda's study found a smaller difference between groups on exposed PVs (Cohen's d = 0.32, a small effect

size). This result implies that a conceptual approach might help learners more easily guess the meanings of unknown PVs.

Another study that used a very similar approach is my study (Shiiba, 2021). The participants were 20 university English as a second language (ESL) learners split evenly into experimental and control groups. The control group were simply told to memorize 35 PVs during the 10-minute period. The PVs were listed alphabetically with their English meanings. On the other hand, the experimental group were instructed on the metaphorical extensions of eight particles. This study contained more PVs than Yasuda's (2010), but experimental participants were given instruction on PVs plus 10 minutes of time to memorize. Both groups were given a gap-filling post-test with 30 questions where they had to choose the correct particle. Ten items were non-instructed verbs using prepositions with prototypical meanings (such as '*go down* a hill'), 10 were instructed PVs with metaphorical meanings, and the other 10 were non-instructed metaphorical PVs. The overall results showed that there was no statistical difference between the two groups on any of the types of PVs (experimental and control), indicating that the study did not find a significant effect of instruction on PVs with metaphorical meanings. However, the study did find that the experimental group outperformed the control group on non-instructed PVs with prototypical meanings, suggesting that instruction on metaphorical extensions of particles may have indirectly

improved overall knowledge of particles. It is worth noting that there are some limitations to these studies. Firstly, the sample sizes were relatively small, which may limit the generalizability of the findings. Secondly, the studies only examined immediate effects of instruction and did not assess long-term retention. Finally, the studies focused solely on particle verbs in isolation and did not examine the use of these verbs in context or in communicative situations.

One might speculate on the reasons for not finding a difference between groups in the study; a larger number of PVs may have led to more confusion among the learners and a smaller number of participants may have not provided enough foundation to find true differences. I do not argue that the results of this small study nullify the findings by Yasuda (2010). However, while conducting the experiment (Shiiba, 2021), I discovered that the method of teaching PVs based on metaphorical extensions was not broadly generalizable. Specifically, it is not possible for all PVs to extend neatly from specific literal definitions to a constricted number of metaphorical ones. For example, Yasuda (2010) taught participants that "removing/excluding is out" (as in the PVs *leave out* and *rule out*) and "searching and finding is out" (as in *figure out* and *make out*). However, even though Yasuda listed the PV with *out* as fitting in the two categories, this idea for the metaphorical meaning of *out* would not be applicable to the meaning of *carry out*: "Perform or complete" (Garnier & Schmitt, 2014, in PHaVE List, p. 9).

Therefore, a new extended definition would be necessary to explain the metaphorical meaning of ‘out’; in creating a list of PVs to use in the experiment I saw that learners would have to repeat many times to understand all of the meanings of PVs, “which may pose a considerable demand on memory and mute the beneficial effect of trying to find a way to remember metaphorical meanings of PVs” (Shiiba, 2021, p. 7). Thus, it could be that the above cognitive linguistic approach is not appropriate for immediate result in knowledge. Therefore, although a cognitive approach is one way that has been proposed for helping LX users study PVs, it has not yet been shown to be an ideal and efficient learning approach for a large number of PVs (for the term ‘LX,’ see Footnote 1).

In conclusion, the acquisition of PVs is an important aspect of language learning, but it can be challenging for language learners due to their idiomatic nature of PVs. While the cognitive linguistic approach has been proposed as an effective approach for teaching PVs with metaphorical meanings, it may not be appropriate or efficient for all types of PVs. Further research is needed to explore the effectiveness of different approaches to teaching PVs, and to identify the most effective methods for helping language learners acquire and use PVs in context.

Polysemy

The investigation of metaphorical extensions of PVs in previous research was motivated

by the recognition of their polysemy. Polysemy refers to the situation where words have multiple meanings, which can pose challenges for language learners who often resort to consulting a dictionary in order to ascertain the correct usage.

In a study conducted by Jorgensen (1990), nine university students were asked to identify the various meanings of six highly polysemous nouns: *head*, *life*, *world*, *way*, *side*, and *hand* - which have over eleven meanings listed in a dictionary. The participants were tasked with a sorting activity, yet they struggled to identify more than three meanings. This suggests that the presence of polysemous words may hinder the acquisition of vocabulary by LX learners, as they grapple with distinguishing the correct meaning of a word from among its multiple meanings.

The polysemous words have both core and non-core senses. In general, the core meaning of a word is based on concrete concepts. For example, the core meaning of *head* is "the top part of your body that has your face on the front and is supported by your neck," while the metaphorical senses include "a leader or person in charge of a group or organization" and "the front or the most important position" (Pearson Education Limited, n.d.).

PVs are also highly polysemous, and the literature debates whether PV meanings should be treated with different dependencies or whether they are prototypical or metaphorical (Gardner & Davies, 2007; White, 2012). PVs have prototypical/literal and

metaphorical/figurative meanings that can range from transparent to opaque (Celce-Murcia & Larsen-Freeman, 1999; Moon, 1997). The prototypical meaning of a PV is defined as the situation in which it has a spatial/literal sense, while the metaphorical meaning is extended from the literal meaning (Lakoff, 1987; Kurtyka, 2001; Rudzka-Ostyn, 2003; Tyler & Evans, 2003). For instance, the meaning of *go down* is literal when it refers to moving to a lower level or position, while it is metaphorical when it refers to decreasing in value or amount (Garnier & Schmitt, 2014). The prototypical meanings of PVs should be easier to understand and employ than their metaphorical meanings due to the features of literal meanings rooted within the spatial senses (Yasuda, 2010). However, English PVs with opaque meanings may pose challenges to learners of English in acquiring them.

Phrasal Verb Lists

It is a well-established fact that non-native speakers of a language have less knowledge of PVs compared to native speakers. However, the question of which PVs should be incorporated into a vocabulary learning program remains a point of debate.

Various lists of PVs based on corpus analysis have been created. Gardner and Davies (2007) produced a list of the top 100 common PVs based on the analysis of *British National Corpus* (BNC), which found that the top 20 lexical verbs found in PV

constructions (e.g., *go*, *come*, and *take*) account for 53.7% of all the PVs in the BNC.

Another list, created by Liu (2011), expanded the scope by incorporating data from both the BNC and the Brigham Young University (BYU) *Corpus of Contemporary American English* (COCA) (Davies, 2008-) across various registers (i.e., spoken, fictions, magazines, newspaper, and academic writing). Liu revealed that there were differences in the frequencies of PVs across these registers. PVs were observed o be less frequently found in writing-oriented registers, such as magazines and newspapers, and particularly infrequent in academic writing. However, it should be noted that Liu's list did not take into account the polysemy of PVs.

In an effort to address the polysemy issue, Garnier and Schmitt (2014) developed the PHaVE List which comprises the 150 most frequent PVs. To determine the polysemous meanings of PVs, Garnier and Schmitt conducted a corpus search and identified the frequent senses that accounted for over 70% of the occurrences. However, due to the limited number of PVs with such high frequencies, additional meanings were included if their frequency was at least 10%, along with the corresponding percentages for each meaning. This information can help learners and teachers prioritize which meanings to focus on, although less common senses are not negligible. For this study, only the most common meanings of PVs were selected. This is the first step in identifying which meanings of PVs are most meaningful to implement into an actual

vocabulary program.

It is worth noting that there may be dialectal variations in the usage of PVs. For instance, a corpus analysis comparing American English and British English PV usage found there are differences in the frequencies of PVs between the BNC and the COCA (Liu, 2011). Although Liu identified an overall similarity in PV usage, there were significant frequency differences between the two corpora. The observed frequency of the 150 PVs in COCA is 1,424,836, while in the BNC, it is 322,517. However, it is important to consider that the COCA covers a more recent time span, starting from 1990 to 2020, while the BNC covers data from the 1980s to 1993. These temporal differences may contribute to variations in PV usage between American and British English. Additionally, it is interesting to note that specific PVs may exhibit distinct usage patterns in different varieties of English. For example, Liu (2011) explains that the disparity in frequency of *shut down* between the BNC and COCA can be attributed to the distinct usage patterns in British English. In British English, the meaning of temporarily closing a business is often expressed using *shut up*, whereas in American English, this meaning is almost always conveyed by the phrase *shut down*. These variations in usage highlight the importance of considering not only the overall frequencies but also the specific contexts and meanings associated with PVs in different dialects. Therefore, when selecting PVs for a vocabulary learning program, it may be

beneficial to take into consideration both the commonalities and potential dialectal differences in PV usage.

Although subsequent studies have conducted corpus analyses using the PHaVE List (e.g., Akbary et al., 2016; Alangari et al., 2019), there is still a gap in research regarding the integration of the PHaVE List into an actual vocabulary learning program. This study aims to address this gap by incorporating the PHaVE List into an innovative vocabulary program.

Avoidance of Phrasal Verbs by Learners

The learning process for LX learners is further complicated by the presence of multiple meanings associated with PVs, which often leads to their avoidance of PV usage. This tendency to avoid PVs is indicative of the proficiency and fluency gap that exists between native speakers and non-native learners in storing and utilizing formulaic sequences (FSs) (Conklin & Schmitt, 2012; Wray, 2000). Consequently, LX learners face a significant challenge in acquiring competence in this specific aspect of English language acquisition. It is worth noting that empirical studies have provided evidence highlighting the influence of L1-English differences on the avoidance behavior exhibited by learners in relation to PV usage.

For instance, Dagut and Laufer (1985) conducted a study to investigate the frequency of avoidance of literal, figurative, and completive PVs by Hebrew English as a Foreign Language (EFL) university learners. One-hundred eighty students in total were given one from three different tests: multiple-choice test, verb translation test, and verb memorizing test. The results indicated that the Hebrew learners avoided using figurative PVs and preferred one-word verbs. The researchers argued that the avoidance was due to the significant structural and typological differences between Hebrew and English, rather than intralingual factors. In Hebrew, prepositions are expressed as one-letter prefixes, and phrases can consist of one word, unlike in English.

Similarly, Liao and Fukuya (2004) investigated the avoidance of figurative PVs by Chinese EFL learners of English. The study compared three types of tests: multiple choice, translation, and recall. The tests consisted of 15 dialogues, each containing 15 pairs of PVs and single verbs. In the multiple-choice test, four options were provided, one of which was a correct PV, one was an equivalent single verb, and two were distractors consisting of one PV and one single verb. In the translation test, the verbs were omitted from the 15 dialogues and their Chinese translations were provided at the end of each dialogue. Students were asked to translate missing verbs from Chinese to English. The recall test included 5 different dialogues as distractors in addition to the same 15 dialogues written with PVs, and learners were required to remember them for

10 minutes. They returned after one hour were asked to fill in the missing verbs in the 15 dialogues. Seventy Chinese university students participated in the study and were randomly assigned to one of the three different 10-minute tests. Of the 70 participants, 25 were advanced learners and 45 were intermediate learners. The multiple-choice test was taken by 10 advanced and 15 intermediate learners, the translation test was taken by 10 advanced and 15 intermediate learners, and the recall test was taken by 10 advanced and 10 intermediate learners. The findings of Liao and Fukuya's (2004) study indicated that learners who took the translation test were more likely to avoid using figurative PVs (multiple-choice: $M = 0.73$ for advanced, $M = 0.43$ for intermediate learners, translation: $M = 0.35$ for advanced, $M = 0.26$ for intermediate, recall: $M = 0.76$ for advanced, $M = 0.48$ for intermediate) than literal PVs (multiple-choice: $M = 0.83$ for advanced, $M = 0.50$ for intermediate learners, translation: $M = 0.83$ for advanced, $M = 0.70$ for intermediate, recall: $M = 0.83$ for advanced, $M = 0.55$ for intermediate). Notably, the advanced Chinese graduate students did not exhibit a tendency to avoid the use of figurative PVs, despite the fact that the Chinese language lacks the same structure as English PVs. The advanced learners had been exposed to an English-spoken environment for more than 9 months, while the majority of intermediate undergraduate students had no experience of living overseas. Liao and Fukuya (2004) suggested that the experience factor influence the avoidance of using PVs.

Moreover, there are typological differences between learners' L1s and English that affect the acquisition of PVs. Japanese, Korean, Turkish, Spanish, and Hebrew have the characteristics of the structure motion + path and are known as 'verb-framed languages' (V-languages). V-languages denote paths via the verb itself, and are not mediated by a particle (Talmy, 2000; Cadierno, 2008). Therefore, Japanese EFL learners may assume that English PVs are inseparable words in the same way as they are in their L1s and that the element of particles is arbitrary in PVs. Compared to V-languages, English, German and Chinese are considered to be 'satellite-framed languages' (S-languages) (Talmy, 2000). The core schema of the path trajectory in S-languages can be expressed by a particle without a verb. As Slobin (2005) pointed out, S-languages are more likely to have descriptions of path trajectories than are V-languages. For example, English PVs are often translated in verb-verb compounds in Japanese (Taniwaki & Tono, 2009). *Cut down* would be translated to *kiri-taosu* 'cut-topple' (p. 320). This difference in the nature of V-languages and S-languages may be one of the main obstacles for learning PVs. The differences between learners' L1 and English causes inevitable challenges that they need to overcome to master English PVs.

Knowledge of Learners' Phrasal Verbs

Several studies which tested English LX users' knowledge of PVs have found that it is

rather poor (Garnier & Schmitt, 2016; Schmitt & Redwood, 2011). In a study by Garnier and Schmitt (2016), 128 Chilean BA students majoring in English Language and Literature or Teaching English as a Foreign Language (TEFL) in two universities were given 100 productive questions on a gap-fill test to measure their knowledge of PVs. The PVs on the test were 40 polysemous PVs containing 100 meanings. The list was based on the *Corpus of Contemporary American English* (COCA) (Davies, 2008-), with frequency counts that reflected their corpus analysis, and the items in the PHaVE List were used; all the 40 PVs had the first and second most frequent meanings, 17 out of the 40 PVs had the third most frequent meanings, and 3 out of the 40 PVs had the fourth most frequent meanings. The result showed that the students answered the gap test correctly 44.5% of the time for the most frequent meaning, 40.1% for the second, 31.6% for the third, and 44.3 % for the fourth most frequent meaning on average. In other words, there was not much difference between knowledge of the more and less frequent meanings of PVs. Additionally, the percentages of PV knowledge were less than 50% even for the PVs with the most frequent meanings. This is surprising since learners of English have PV knowledge of less than 50% even with the PVs that have the most frequent meaning. This may imply that learners prefer to use lexical items rather than PVs because they are more explicitly taught in their school. The acquisition of PVs thus seems to be very difficult for learners if they do not use any specific

learning approach. To date, no study has focused on how PVs should be learned and how many PVs could be learned if PVs were introduced into a vocabulary learning program.

Based on the findings from studies assessing English LX users' knowledge of PVs and the observed difficulties in acquiring PVs without specific learning approaches, there is a clear need to investigate strategies that can enhance the acquisition and retention of PVs.

In Japan, the curriculum for English education at junior and senior high school is often based on exams to enter higher education, which prioritize the teaching of individual lexical items (e.g., vocabulary words and grammar structures) that are more commonly assessed in these exams. These exams tend to focus on discrete vocabulary and grammar knowledge rather than the ability to use FSs, including PVs, in context. Therefore, it can be assumed that, like the previous study on Chilean university students, Japanese university students may not have extensive knowledge about PVs.

Memory and Spaced Retrieval Practice

Without retrieval, memory attenuates and will soon be forgotten if not reinforced (Ebbinghaus, 1993). One of the important questions in LX vocabulary acquisition is the issue of how memory works and how best to learn a large number of words. To transfer

vocabulary items from more short-term memory into long-term memory, repeated retrieval practices are indispensable (Kornell, 2009). After learners encounter and learn a new word, for example, they are asked to use the learned word through some activity, which requires them to retrieve information from their memory. This learning activity is an example of retrieval practice. Another important question in LX vocabulary acquisition is *when* to practice retrieval in order to be conducive to long-term retention. A number of studies have shown that retrieval practice that is conducive to long-term retention consists of repeated learning, forgetting, and relearning (e.g., Cepeda et al., 2006; Cull, 2000; Kang et al., 2014; Karpicke & Roediger, 2007; Logan & Balota, 2008; Nakata, 2015). Karpicke and Blunt (2011) note that retrieval practice is more promising than other activities like massed learning or elaborative concept mapping. Eighty undergraduate students learned a text about science in four learning conditions: study-once group, repeated study group, elaborative concept mapping group, and retrieval practice group. The results from the one-week delayed post-test involving verbatim and inference questions showed that the group conditioned by retrieval practice outperformed better than the other three groups.

Crammed learning, or massed learning, is popular in Japan to prepare for term exams at school or even entrance exams for higher education (Weisman, 1992). This is due to the Japanese educational model that students have been told to just memorize,

promoting passivity. My own anecdotal knowledge shows that students are not able to recall the contents even a few days after memorization or massed practice.

Several studies have shown that a number of learners believe massed study is more effective than spaced study (Kornell & Bjork, 2008; Simon & Bjork, 2001). However, to retain the memory for a long time, retrieval practice with spaced intervals is essential (Nakata, 2017). The famous 'forgetting curve,' which was presented by Ebbinghaus (1993), suggests that memory retention would be reduced so quickly in the short term. Although his empirical research was based on a single-experiment study, the produced result has been gradually accepted, showing that memory attenuates in only one hour, forgetting 75% after the first recall, if information is not reinforced. However, there is a fact that review affects memory. An initial review enables learners to retain the memory in the short-term. A second, third, fourth reminders at different intervals are helpful to their retention in the long run.

Psychological studies have also shown that a spaced repetition system is a better way to learn more vocabulary (Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007; Kornell, 2009; Pyc & Rawson, 2007). Learning with interleaved delay, which is called spaced learning, is more effective than massed learning (Janiszewski et al., 2003; Kornell, 2009). This is also applicable to vocabulary acquisition. Vocabulary that is learned with interleaved delay can be recalled in the long run. Spaced learning can

contribute to memory retention in terms of LX vocabulary (Nakata, 2015).

According to Karpicke and Bauernschmidt (2011), spaced learning comprises some different types of measurements such as absolute and relative spacings. Absolute spacing measures the total time of the intervals between study hours, while relative spacing is categorized into three types of learning schedules based on the intervals between study hours: expanding spacing, equal spacing, and contracting spacing. Recent studies have shown that there are no differences between the effect of expanding and equal spacing on LX vocabulary acquisition (Pyc & Rawson, 2007; Karpicke & Bauernschmidt, 2011; Kang et al., 2014; Nakata, 2015; Kanayama & Kasahara, 2016). Nation (2013) also claims that expanding spacing is equally effective to equal spacing. In fact, according to several studies, on one hand, expanding spacing enables learners to retrieve from their primary memory which can contribute to short-term memory because the interval between the first encounter and the first retrieval is shorter than that of equal spacing (Karpicke & Roediger, 2007; Storm et al., 2010). On the other hand, equal spacing would be conducive to retention of memory as retrieval can be conducted from long-term memory. Considering the fact that recent flashcard programs, such as *Quizlet* and *Anki*, are capable of judging automatically which word should be prioritized according to the individual remembering rate, it seems that the types of spaced learning are less relevant for this study.

There is one study that specifically investigated the learning conditions of massed-repetition and spaced-repetition for vocabulary learning. In the study conducted by Kornell (2009), the focus was on comparing the effectiveness of massed-repetition learning and spaced-repetition learning conditions, using flashcards with varying intervals. Twenty-five university students in California studied 40-word pairs in total of a target word and its synonym (e.g., *effulgent*: *brilliant*) for four days. Each group studied different 20-word pairs that were presented on computers. Students in massed learning condition learned 5 out of the 20-word pairs eight times in one day and this process was repeated for four days. Learners using spaced repetition studied the 20-word pairs two times per day for four days. The cued-recall post-test result on the fifth day showed that the percentage of questions answered correctly for the spaced learning condition was 54 percent whilst the massed learning group answered only 21 percent correctly (Cohen's $d = 1.18$).

There is, however, one shortcoming in this study: Although Kornell's (2009) results support the fact that spaced repetition is strongly effective on learning vocabulary, the learning duration was only four days. Thus, Kornell's study brought up a question of how much words the students remember in the long run. It seems that no research has yet studied whether spaced learning would assist in learning English PVs in the long term. Therefore, the aim of this study is to investigate whether spaced

learning will have more positive effects on learning and recalling English PVs than massed learning on a long-term scale. To achieve this goal, spaced learning is employed for students to learn PVs.

Flashcard Learning

The use of flashcards is now a ubiquitous learning strategy and enables learners' self-regulated study (Kornell & Bjork, 2007). Flashcard learning has been regarded as an effective vocabulary learning method if spaced repetition is used (Baddeley, 1997; Kornell, 2009). Moreover, several studies have regarded flashcard learning as an essential learning activity for enabling learners to remember a number of vocabulary words in short-term studies (Fitzpatrick et al., 2008; Nation, 2001). For example, Fitzpatrick et al. (2008) featured a 41-year-old woman teaching linguistics at a university who was asked to learn 300 high frequency Arabic vocabulary words over 20 days and to spend a maximum of 30 minutes on learning 15 words per day. After 20 learning sessions, two different tests were conducted four times: a recall test, in which she was asked to provide Arabic translations with English cues, and a recognition test asking her to translate English words into Arabic words. The first set of tests was given immediately after the last learning session (283 correct responses for recall test and 286 for recognition test), the second tests were provided two weeks after the first tests (191

and 262), and the third tests were conducted four weeks after the second tests (135 and 221). The results of the fourth tests, which was administered 10 weeks after the last learning session, show that the subject still performed well on recall and recognition and tasks, with 149 and 219 correct responses respectively for recall and recognition

Nakata (2008) suggests that the use of flashcards makes it possible to classify words that are already learned and those that are not learned yet. He examined three different types of vocabulary learning based on word lists, printed cards, and computers by 217 senior high school students. Students in the list and card group were allowed to write freely on sheets of paper to memorize words while those in the computer group learned words by typing a correct English word corresponding to its Japanese translation. The stack had 10 English vocabulary words with their Japanese translations. The four-day delayed recall post-test indicated that learners studying words through computers outperformed the list use condition (Cohen's $d = 0.52$) while there was no difference between the computer use group and the card use group ($d = 0.26$) and the list and card groups ($d = 0$ .27). This suggests that productive retrieval practice after learning words receptively was the most effective learning. This is in consistent with Karpicke and Blunt's (2011) study on concept mapping in the context of science learning.

If experimental participants are going to undertake spaced repetition retrieval,

the next question is how often they should perform the repetition. The exact number of repetitions for learning new vocabulary is still controversial. Nation and Wang (1999, p. 363) note that "[T]here is no set number of repetitions that will ensure learning." Nation (2001) overviewed a number of studies and claimed that roughly 5 to 20 repetitions were required to acquire any vocabulary word. Moreover, Brown et al. (2008) showed that picking up a new word from listening-only requires 20 encounters at least, but this was not sufficient to acquire some word meanings, while 50 or higher encounters may be required when doing graded reading. Nevertheless, "[D]isregarding the exact number of repetitions required, the important point is that recycling is necessary, and if it is neglected, many partially learned words will be forgotten, wasting all the effort already put into learning them" (Nation, 1990, p. 45).

In setting up an actual vocabulary program, program directors face an additional decision on how many words students need to study every week. It is also still not clear how many words should be presented in one stack. Several studies have shown that a larger block size is more effective than a small one (Crothers & Suppes, 1967; Kornell, 2009). According to Kornell (2009), who defined 20 words in one stack as a large block size, a spaced condition where learners studied one large stack of 20 words twice during each of four sessions outperformed the massed condition where learners studied small stacks of 5 words eight times each. The results were obtained by an online test 24 hours

after the memorization. Clearly, the field has yet to really investigate this question of an optimal block size in a way that applies to long-term vocabulary learning. Nation (1982) claims that the use of a large block size is effective when task difficulty is low. Thus, a large block size was used in this study, which will be more practical for pedagogical application. One block had 25-50 words including PVs and academic vocabulary words.

2.2 Research Questions

This study aimed to investigate two research questions:

1. How can a study program be designed to enhance learning outcomes for PVs?
2. To what extent can learners acquire PVs over one academic semester?

2.3 Methodology

Participants

The target participants for this study were 143 first-year students who were majoring in English language at a university in the Kitakyushu area of Japan.[2] Students who did not take either the PV pre-test or post-test or did not agree to participate by not responding to the informed consent form through *Google Forms* were not included in the study. In

[2] The *hensachi* of the university department is ranked at between 50.0-52.5 The Common European Framework of Reference for Languages (CEFR) level is between B1-B2.

total, 125 students participated in the first semester, while 103 students participated in the second semester.

While previous research has suggested that learning with spaced repetition is superior to massed learning, it may not always be practical to include a control group to establish causality. Nonetheless, it is worth noting that numerous studies have consistently shown the benefits of retrieval practice, which involves repeated learning, forgetting, and relearning, for enhancing long-term retention (e.g., Kang et al., 2014; Karpicke & Roediger, 2007; Logan & Balota, 2008; Nakata, 2015). Considering these findings, it was decided not to include a control group in the present study.

Learning Program

The study was conducted over the course of one academic year, which is divided into two semesters of 15 weeks each. In the first semester, students were given 12 weeks to learn a total of 60 PVs and 540 academic vocabulary words. In the second semester, students had to learn another 60 PVs and 300 academic vocabulary words within the same 12-week period. The use of PVs in the study was derived from the PHaVE List. Additionally, academic vocabulary words used in the university's vocabulary learning program were obtained from the Academic Vocabulary List (AVL), developed by Davies and Gardner (2013).

Learning academic vocabulary was a part of the university learning program, in addition to learning PVs. The AVL is a comprehensive list of words that are essential for academic learning. These words are grouped by families and include both academic and technical vocabulary that are commonly used in specific domains. The AVL contains a total of 1991 word families, and each word family has a frequency count listed beside it. In this study, only academic vocabulary, not technical vocabulary, was selected from the AVL and ordered by frequency. The top 840 academic vocabulary were then used for one academic year. However, it is worth noting that this study focused on learning PVs and did not analyze learners' outcomes on academic vocabulary.

To motivate students to engage in self-directed learning of the vocabulary, they were required to take weekly review tests in class. In order to pass each test, students had to score at least 80 percent on eight out of twelve tests. This provided an incentive for students to continue studying and reviewing the material throughout the semester.

Flashcard Learning

To implement the flashcard learning program, the *Quizlet* platform was used in this study. All students had access to the same vocabulary learning tool, which could be used on both smartphones and computers. The students were given free year-long

accounts with *Quizlet Plus*, which enabled them to study vocabulary using spaced repetition (the Leitner system).[3] Spaced repetition ensured that incorrectly answered cards would appear more frequently than correctly answered cards.

The vocabulary stacks in *Quizlet* were standardized for all students, with 45 academic words and 5 PVs per stack in the first semester, and 20 academic words and 5 PVs per stack in the second semester. In the second semester, the number of academic vocabulary words was reduced from 45 to 25 words in response to student requests.

Participants were encouraged to practice the vocabulary for 20 minutes a day, five days a week but no verification of this schedule was performed. Each stack appeared weekly on *Quizlet* and became visible for the students.

In the *Quizlet* stacks, Japanese translations were given for both the AVL words and the PVs. I made the Japanese translations for the first 500 AVL words provided by Crandell (2017) as well as for the remaining words. Each card in the stack consisted of a

[3] *Quizlet Plus* is a paid application. *Quizlet* has since discontinued their spaced repetition program. This feature was only available if students used a computer to review words and typed in the answers, so not all students might have availed themselves of this feature since studying vocabulary on their mobile devices was also possible. One program which does provide spaced repetition with mobile devices is *Anki*, but it takes a little more work than *Quizlet* to install and set up. In addition, technical burden that would reduce the risk of student participation could be considered, so I chose the *Quizlet* system.

target English word on one side of a card and the meaning in Japanese on the other side.

Weekly vocabulary quizzes were given to students through *Google Forms* in the first semester and through *Moodle*, which is an online learning system at university, in the second semester. The reason why the testing method was changed is that *Moodle* was able to randomize questions so every student had a different test. The quizzes were conducted weekly in class under the supervision of the class teachers and administered through *Google Forms* or *Moodle*. The format of the quiz was a translation from Japanese words into English, what Stoeckel et al. (2019) calls a 'form recall' test.

To ensure consistency in grading, potential misspellings were marked as correct if they were understandable to native speakers, and some flexibility was allowed to account for vocabulary not included in the AVL. For example, misspellings such as *futhermore* or *futhurmore* for *furthermore* and the flexibility such as *promising* or *feasible* for *potential* (可能性のある) are regarded as correct. Two judges, including a native speaker of English, marked each test for interrater reliability, and any disagreements were resolved through discussion. Overall, these measures were taken to ensure that all students were tested on the same sets of words and to maintain consistency in scoring.

The quiz had a time limit of 10 minutes and included previously learned words to encourage review. The vocabulary was chosen randomly from the stacks on *Quizlet*.

Students received feedback on their quiz performance a few days after taking the quiz.

Phrasal Verbs

This study used the most frequent 120 PVs from the PHaVE List. Sixty PVs were selected from the top 120 PVs for the first semester (#1 to #60) and the remaining 60 PVs (#61 to #120) were used for the second semester. Only the most frequent meaning in the list was used. Five PVs were chosen for weekly learning in order not to overlap with PVs having a similar meaning in the same week (e.g., *go in*, *get in*, and *come in*). Moreover, in order to avoid the bias of higher and lower frequencies in the weekly five PVs, the 120 PVs were divided by 24, which made 5 sections, and one PV was selected from each section for weekly use.

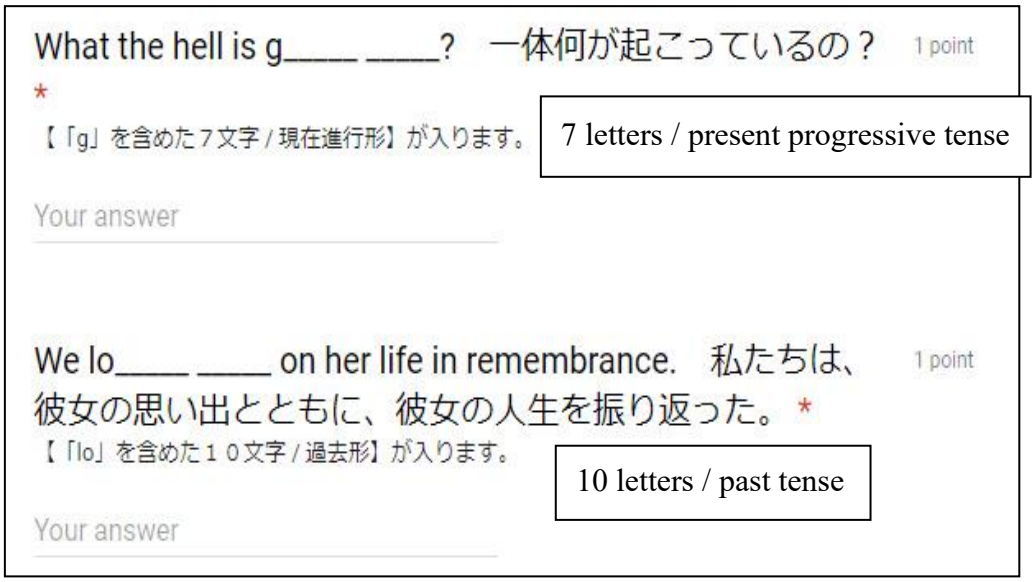
Tests

Before the experiment, all students took a pre-test which aimed to measure their prior knowledge of PVs. The pre-test was administered using *Google Form* and consisted of a fill-in-the-blank task. A total of 50 PVs were included in the test, selected from the targeted 60 PVs for the semester.

In order to reduce the impact of context, which could make non-target PVs or expressions valid answers, the first two letters of the PV and a Japanese translation were provided as hints. However, if the original verb of a PV was two or three letters long

(e.g., *go*, *get*, and *put*), then only the first letter of the verb was provided as a hint along with the Japanese translation (e.g., *go out*: "g___ ____" for the present tense, "w____ ____" for the past tense). The number of words and the verb tense in Japanese to be filled in were also shown above the English sentence, based on the feedback received during pilot testing with native and non-native speakers (Figure 1).

Figure 1. *Examples of Phrasal Verb Gap-Filling Test*



A post-test was administered at the end of each academic semester to assess the students' overall learning development. The post-test was also administered in *Google Forms* and consisted of the same fill-in-the-blank task as the pre-test, using sentences from PHaVE Appendix A. The order of the questions was randomized to reduce the potential impact of memorization. To ensure that the students were not simply recalling

the answers from the pre-test, they were not given access to the correct answers for the pre-test. The tests were designed to be the same as weekly quizzes in terms of requiring students to translate Japanese into English.

The decision to use a productive test for PVs was based on two primary reasons. First, it allowed for a direct comparison of results to previous studies that used similar measures, such as the test used by Schmitt and Redwood (2011). Second, a multiple-choice test for PVs would have presented challenges due to their polysemous nature. Without translations, it would be difficult to ensure that the test focused solely on the target PV. However, if translations were included, it would be too easy for learners to guess the correct answer. A productive test, on the other hand, requires learners to produce the PV in context, which is a more accurate measure of their understanding and use of the target item.

2.3.1 Results

Before analyzing the data, statistical analyses using paired samples *t*-tests were performed to examine normality. Normal distribution was found in gain scores of the pre-test and post-test. An alpha level of .05 was used to show whether there is a statistical significance. Then, paired samples *t*-tests were performed to determine whether there were statistically significant differences between two groups. Effect sizes

were also reported to look at how much difference were between the pre-tests and post-tests.

1st Semester

After a 4-month study on PVs, the results presented in Table 1 indicate that the mean scores of all students on a challenging production test almost doubled. This was expected as students had to study PVs during the semester. Descriptive statistics reveal a significant difference between pre-test (M = 15.98, SD = 7.10) and post-test (M = 35.26, SD = 8.29). A paired samples t-test on the gain scores found that the difference was statistical, $t(124) = 23.06$, $p < .00001$. More importantly, a Cohen's d effect size for the gain score was found to be $d = 2.50$, indicating that the increase in mean score was more than two times the standard deviation of the students. This indicates that the students as a whole made good progress in producing appropriate PVs in sentences, which was the goal of the learning program.

Table 1. *1st-Semester Students' Phrasal Verb Pre-Test and Post-Test Scores*

Tests	N	Minimum	Maximum (Max. 50)	Mean	SD	95% CI	
						Lower	Upper
Pre-test	125	0	31	15.98	7.10	14.73	17.24
Post-test	125	1	49	35.26	8.29	3.80	36.73

Gain scores	0	45	19.28	9.35	17.62	20.94

The 95% confidence interval (CI) in Table 1 for the gain score does not pass through zero, indicating a statistical difference, and an actual gain in scores lies somewhere between 17.62 and 20.94 with 95% confidence.

Moreover, to assess the internal consistency and reliability of the measures, the Kuder and Richardson Formula 20 (KR20) test was employed for both the pre-test and post-test data. The KR20 coefficient was calculated to be 0.994 for the pre-test and 0.993 for the post-test, indicating high internal consistency and reliability of the scores on both tests between two raters.

Figure 2 shows the correlation between pre-test and post-test scores for all 125 students, with most students falling above a 45-degree diagonal line. This suggests significant progress on the test. Therefore, spaced repetition could have a great effect on the students' learning outcome of PVs overall.

In conclusion, the results of this study show that the implementation of spaced repetition was effective in helping students improve their ability to produce PVs in sentences. The significant increase in mean scores on the post-test compared to the pre-test, coupled with a large effect size and a 95% CI that did not pass through zero, indicates that the gains were statistically significant. Furthermore, the high internal consistency and reliability of the measures used in the pre-test and post-test data suggest

that the results are robust and can be reliable.

Figure 2. *Scatterplot of Gain Scores on Phrasal Verb Pre-Test and Post-Test in 1st Semester*

These findings have important implications for language teachers and learners who are seeking effective ways to improve their ability to use PVs in a meaningful way. Spaced repetition, as demonstrated in this study, can be a useful tool in achieving this goal. Future research could explore the potential benefits of spaced repetition in other areas of language learning and investigate the optimal spacing intervals for repeated exposure to target language items.

2nd Semester

Table 2 presents the results of the pre-test and post-test administered during the second semester. These tests covered 50 out of the 60 PVs targeted for the second semester, which were different from those covered in the first semester. The mean score increased by approximately 7 points, which represented less progress than was achieved in the first semester. It is interesting to note that the mean score on the pre-test was nearly10 points higher than it was in the first semester, although the post-test scores were similar. A paired samples *t*-test on the gain score found a statistically significant difference, $t(102) = 7.69$, $p < .00001$. The Cohen's *d* effect size for the gain score was $d = 0.85$, which is still a large effect size. This suggests that most students were able to employ the appropriate PVs in more than 50% of the items, similar to the result from the first semester.

Table 2. *2nd-Semester Students' Phrasal Verb Pre-Test and Post-Test Scores*

Tests	N	Minimum	Maximum (Max. 50)	Mean	SD	95% CI	
						Lower	Upper
Pre-test	103	7	39	25.44	6.60	24.15	26.73
Post-test	103	11	47	32.03	8.85	30.30	33.76
Gain scores		-22	31	6.59	8.70	-8.29	-4.89

In order to evaluate the internal consistency and reliability of the measurements, the Kuder and Richardson Formula 20 (KR20) test was employed for both the pre-test and post-test data. The KR20 coefficient yielded values of 0.987 for the pre-test and 0.994 for the post-test, signifying a substantial level of internal consistency and reliability between the two raters' scores on both assessments.

Figure 3. *Scatterplot of Gain Scores on Phrasal Verb Pre-Test and Post-Test in 2nd Semester*

Figure 3 illustrates the correlation between the pre-test and post-test scores for 103 students. The graphical representation shows that most students made good progress on the test, although more students fell below a 45-degree diagonal line than in the first semester result. This may indicate that students were less motivated during the second semester than in the first semester, as fewer students were able to attain the 80% vocabulary score on the weekly tests.

2.3.2 Conclusion

In this study, the effectiveness of flashcard learning with retrieval practices was investigated as a study program for learning PVs. The results of this research have practical and pedagogical implications, as well as providing insights that can be applied to theoretical issues.

The findings revealed that 35 out of 50 first-semester PVs (70%) and 32 out of 50 second-semester PVs (64%) were correctly answered on challenging gap-filling tests at the end of the semester. The gain scores for the first semester were 19, while for the second semester, they were 7 on average. The use of spaced-retrieval practices with flashcards had a significant positive effect on participants' learning performances on PVs.

Although the gain scores of the second semester were lower, it is worth noting that students had a larger pool of available answers to choose from in the second semester as they had been learning PVs all year long. However, the question of how long their knowledge might last without continued reinforcement remains a concern, as most students may not continue studying the words after the end of the year. Thus, a delayed post-test was conducted one year or six months after students had learned the PVs to examine the long-term retention of the vocabulary.

2.4 Methodology of Delayed Post-test

Participants

The participants in the delayed post-test study were sophomore students who had previously gone through flashcard learning of PVs in their first-year period. Due to COVID-19 prevention measures, all instructions and learning in the participants' sophomore year were conducted remotely through the university's online learning system *Moodle*. Only those students who agreed to participate and provided informed consent via G*oogle Forms* were included in the experiment, resulting in 27 participants being enrolled in this study.

Tests

A delayed post-test was administered on *Moodle* during the second semester to evaluate the participants' long-term retention of the PVs. The test was conducted one year after the students had completed learning 60 PVs in the first semester and six months after they had finished learning 60 PVs in the second semester of their first year. The delayed post-test consisted of 60 PVs, with 30 PVs from each semester. However, 25 out of the 60 first-semester PVs were excluded from the test as they had already been tested more than four times during the students' first year for experimental purposes. This was done to prevent any potential bias caused by overexposure to these PVs.

All the sentences used in the delayed post-test were different from those used in the participants' first-year pre-test and post-test. These sentences were sourced from the PHaVE List, and I provided Japanese translations to all of them. The task type for the delayed post-test was the same as the previous experiment, and the sentences were presented in a sentence completion format. For a complete list of the sentences used in the test, please refer to PHaVE Appendix B.

2.4.1 Results

Prior to data analysis, normality of the delayed post-test scores was examined using one-sample *t*-tests. The results showed normal distribution of the scores. An alpha level

of .05 was used to determine statistical significance. Again, one-sample *t*-tests were performed to examine whether there were statistically significant differences between the scores of the two groups. Effect sizes were also calculated to look at the magnitude of the differences observed.

In order to evaluate the effectiveness of the flashcard learning program, the participants were given a pre-test before the program, an immediate post-test after completing the program, and a delayed post-test one year or six months later. The results of the initial pre-test indicated that participants had a limited knowledge of the PVs, with an average score of 32% (16 out of 50 PVs) for the first semester and 50% (25 out of 50 PVs) for the second semester. Following the flashcard learning program, the participants' immediate post-test scores significantly improved, with an average score of 70% (35 out of 50 PVs) for the first semester and 64% (32 out of 50 PVs) for the second semester. While these results suggest that the flashcard learning program had a positive effect on participants' PV acquisition, the delayed post-test results provided additional insight into the long-term retention of these words.

Table 3. *Delayed Post-Test Scores on Phrasal Verbs*

When learned PVs	N	Minimum	Maximum (Max. 30)	Mean	SD	95% CI	
						Lower	Upper
1st semester	27	10	26	16.37	4.44	14.61	18.13

2nd semester	27	6	23	15.26	4.45	13.50	17.02

Table 3 presents descriptive statistics of the delayed post-test scores. The mean scores for the first-semester and second-semester PVs were 16.37 and 15.26, respectively. These results indicate that, on average, the students retained half of the PVs learned in each semester. Although the mean scores were lower than those of the immediate post-tests (35/50 for the first semester and 32/50 for the second semester), it is still a noteworthy accomplishment given that the students were not expecting the test and had not studied for it. The difference between the scores of the two semesters was not significant, as indicated by a paired t-test ($t = 1.82$, $df = 26$, $p = .08$).

Figure 4. *Distribution of Phrasal Verb Scores on Delayed Post-Test by Semester*

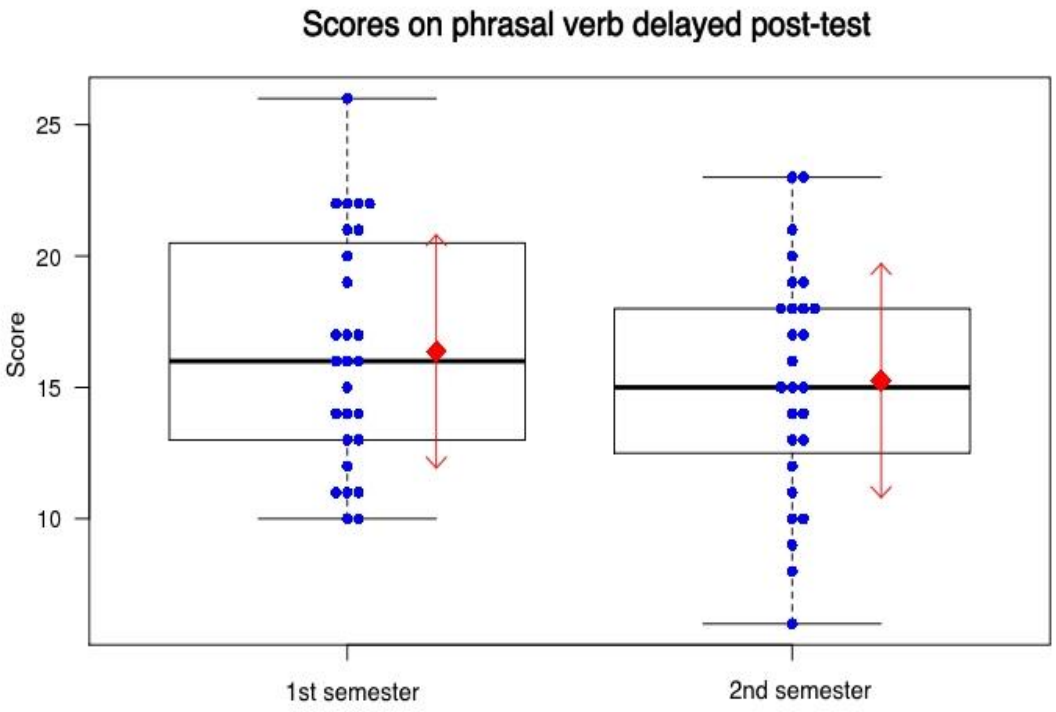


Figure 4 illustrates the distribution of the students' scores for PVs learned in each semester. There was no significant difference between the scores of the two semesters.

In terms of practical implications for implementing a vocabulary program, this study did not determine the exact number of encounters needed for complete learning of the PVs. However, the results show that the students were able to retain approximately half of the PVs they learned over the semesters, even after a one-year delay and without any additional study or reinforcement. It is important to note that the sentences used in the delayed post-test were different from the ones used in the immediate post-tests, indicating that the students were able to generalize their knowledge of the PVs to new contexts.

While the retention rate of approximately 50% is lower than the immediate post-test results, where the students were able to recall roughly 70% of the PVs they had learned, it still represents a significant amount of knowledge that can be drawn upon. The results suggest that a vocabulary program that includes spaced repetition and exposure to PVs in a variety of contexts can be effective in promoting long-term retention of the PVs. However, additional research is needed to determine the optimal number of encounters needed for the complete mastery and retention of PVs.

2.4.2 Conclusion

In the present study, the delayed post-test was conducted one year or six months after students had learned the PVs. The results showed that they still remembered about 50 percent of the PVs. This suggests that the computer flashcard learning method with spaced retrieval practices used in this study was effective in promoting long-term retention of PVs even after the completion of the program.

Notably, the initial pre-test mean scores were 16 out of 50 PVs (32 percent) in the first semester and 25 out of 50 PVs (50 percent) in the second semester. Despite this, a significant number of PVs, namely 15-16 out of the 30 PVs for each semester, were successfully retrieved in the delayed post-test. This finding highlights the utility of the computer flashcard learning method in facilitating not only short-term learning but also

long-term retention of PVs.

Moreover, the students were able to retain a substantial portion of the PVs they had learned even though they were required any mandatory post-program review or reinforcement activities. This underscores the effectiveness of the spaced retrieval method used in this study, as well as the potential for this method in other educational settings. Overall, the results of this study suggest that spaced retrieval practice with computer flashcards can be a valuable tool for facilitating vocabulary learning and retention.

2.5 Discussion

There are several features, significance, and implications to be discussed in the study. The success of the spaced retrieval method used in this study can be attributed to the fact that it is based on the spacing effect, which suggests that learning is more effective when study sessions are spaced out over time. By spacing out the retrieval of PVs, learners were able to retain the words more effectively in their long-term memory.

The findings of the present study are consistent with previous studies that have shown that spaced retrieval practice is an effective technique for promoting learning, not only for vocabulary learning (Goossens et al., 2012; Kornell, 2009; Nakata, 2013, 2015) but also for other domains (Golding et al., 2012; Lu et al., 2021). Specifically, Kornell (2009) found that spaced retrieval practice was more effective than massed

practice, although the present study did not include massed learning condition. While no prior empirical studies have explored whether spaced retrieval practice using flashcards can promote long-term retention of vocabulary, the present study suggests that the use of flashcards with spaced retrieval practice can be an effective strategy for encouraging learners to retain vocabulary over extended periods, such as one year. In contrast to previous studies that utilized immediate or short-delayed post-tests (e.g., Anderson et al., 2009; Fitzpatrick et al., 2008; Kornell, 2009; Nakata, 2008), the present study employed a delayed post-test administered at six months and one year after the immediate post-test. This extended time frame allowed for a more robust assessment of learning and long-term retention.

The findings of this study suggest that language teachers incorporate spaced retrieval practice with computer flashcards into their vocabulary instruction to promote long-term retention of words and phrases. Namaziandost et al. (2020) suggested incorporating spacing into language teaching curricula, instruction, and materials to enhance vocabulary learning in real classroom settings. The present study showed that this approach is useful for teaching vocabulary, specifically PVs, which are polysemous and can be difficult to learn and remember. Additionally, teachers may consider using this method in conjunction with other effective learning strategies, such as elaborative questions, to enhance learning and retention.

While this study provides evidence of the effectiveness of spaced retrieval practice with computer flashcards for promoting long-term retention of PVs, further research is needed to address several limitations of the present study. Firstly, monitoring students' usage of the flashcard program and its correlation with better learning outcomes could offer more detailed insights. The exact number of repetitions needed to learn new vocabulary is still a matter of debate. While Nation and Wang (1999, p. 363) argue that there is no fixed number of repetitions for learning, Laufer and Nation (2012, p. 167) roughly estimate that around ten exposures are necessary for learners to recognize the meaning of a new word. However, Webb (2008) emphasizes the importance of context in understanding word meaning, suggesting that the quality of the context plays a critical role in vocabulary acquisition. Further research is needed to determine the optimal number of exposures or repetitions required for effective vocabulary learning, considering both the quantity and quality of encounters.

Secondly, since the tests were conducted online, it is possible that some participants may have cheated, although there was no penalty for mistakes. It is difficult to prevent cheating in online testing (Munoz & Mackay, 2019). Moreover, the study used a gap-filling test, where participants had to fill in the missing word themselves, based on a given context. Although they were given some clues such as a translation and the number of letters to be filled, the test was challenging for Japanese learners of

English. This type of test measures not only their receptive knowledge of the language, but also their ability to actively produce the target words (Groot, 2000). If the test type had been different, the participants' scores would likely have differed from the results of this study. For example, a test that focuses on meaning recall would probably have yielded different scores. The impact of the test format on scores is dependent on the type of material being tested (Agarwal et al., 2008).

Thirdly, this study did not provide any conclusive findings regarding the optimal number of words that should be learned within an academic semester or the impact of the block size on vocabulary acquisition. Although the participants in the study learned five PVs and 45 academic vocabulary words per week during the first semester and five PVs and 25 academic words weekly during the second semester, it is uncertain whether a larger block size would have resulted in more effective vocabulary learning (Crothers & Suppes, 1967; Kornell, 2009), or if the increased enthusiasm of new first-year students contributed to their success in the study. As a result, further research is needed to determine the most effective block size for long-term vocabulary acquisition.

Finally, in the future, to better control the recall output in a test, I suggest that it may be beneficial to use blanks to show the number of letters in each word or PV and provide a couple of hint letters aside from the first one or two letters of the verb. This format of testing may help learners focus on recalling the spelling of the word, rather

than simply recognizing it based on the first one or two letters of the verb. However, it should be noted that there seems to be no study examining how effective this format is in helping learners focus on recalling the spelling of the word. I believe that it may also allow for a more controlled assessment of their vocabulary knowledge.

Despite these limitations, this study suggests that deliberate study of words can help learners memorize them, and retrieval practices using flashcard learning can contribute to learners' retention of words. Moreover, an interesting area for further research would be to compare the effectiveness of the flashcard learning program on PVs with academic vocabulary words. While the program aimed to teach both types of words, learners may have benefited differently from each category. The recall and recognition of different word classes can vary (San-Mateo-Valdehíta & Chacón-García, 2019). However, little research has directly compared the learning outcomes of multi-word sequences such as PVs and academic vocabulary. Such comparison would provide insights into the relative difficulty of learning these word types and whether distinct strategies are necessary to enhance their acquisition. Furthermore, an interesting avenue for future research would be to explore alternative testing formats that enhance the recall of PVs. For instance, incorporating blanks to indicate the number of letters in each word or PV, along with providing a few hint letters in addition to the first one or two letters of the verb, could potentially improve learners' focus on recalling the

spelling of the word. This format not only encourages active retrieval but also allows for a more controlled assessment of learners' vocabulary knowledge. I was unable to find existing studies examining the effectiveness of this format in promoting spelling recall. Considering the potential benefits it offers, I believe that investigating the impact of this testing format on learners' recall abilities would be a valuable direction for future research in this area.

CHAPTER 3.

Experiment 2: Learning Phrasal Verbs with Tasks

Learning phrasal verbs (PVs) poses a significant challenge for English language learners. One particularly difficult aspect is comprehending the diverse range of meanings that PVs can have in various contexts. Memorizing flashcards can be helpful, but it may not be enough to fully comprehend how to use PVs in context. In fact, Chapter 2 has shown that using flashcards alone may not be enough for learners to fully understand PVs.

Previous studies have claimed that involving context in vocabulary learning is more effective than plain word lists, as will be discussed later. Providing learners opportunities to encounter the appropriate use of PVs in sentences is important and should be promoted with tasks using sentences. On top of flashcard learning, I then came up with an idea that learners could learn PVs using example sentences on *Moodle*, the e-learning system employed at the university. This chapter will thus look into learning approaches for studying English PVs using tasks on the university online learning program.

This chapter aims to investigate the effectiveness of the above-mentioned online additional tasks. Then, by comparing this method with flashcard learning we seek to identify the most effective type of task for promoting PV production ability. The study

also seeks the relationship between task participation frequency, time spent on tasks, and improvement in PV production, with the objective of increasing the number of PVs produced in the production test.

3.1 Literature Review

The Use of Sentences

In considering an ideal and efficient learning of PVs, Chapter 2 showed that flashcard learning with retrieval practices enhanced learners' acquisition pf vocabulary. However, my study found that the use of flashcard learning only would deprive students of the chance of employing the target vocabulary words since they were learning the words without context. On top of flashcards, seeing and using sentences might be important for learners to learn PV use.

Some studies have claimed that involving context in vocabulary learning is more effective than plain word lists. Learners themselves can infer or induce the meaning in sentences (Bialystok, 1983; Nation, 1982). Exemplars can support learners' learning vocabulary and retention in the long run (Baicheng, 2009; Cobb, 1997; Laufer & Shmueli, 1997). The results of Baicheng's (2009) study indicate that learners who engaged in vocabulary learning with example sentences and heavier information processing loads found it necessary to dedicate time to elaborate on the syntactic information of the target items. The inclusion of example sentences in vocabulary presentation serves an important purpose by elevating the information processing load. Baicheng (2009, p.183) also notes that this process can contribute to the storage of information in long-term memory and enhance subsequent retrieval. Furthermore,

Nation (2001, p.231) suggests that target vocabulary should be treated together with an example sentence.

There has been a criticism on the use of decontextualized tasks without enough information on words for learners (Oxford & Crookall, 1990). Yet, Webb's (2007) research suggests that decontextualized tasks are as effective as contextualized tasks. Eighty-four first-year university students in Japan were asked to learn 20-word pairs by matching 20 false words to 20 low frequency vocabulary with their Japanese meanings for 8 minutes through two different conditions: an experimental group ($N = 41$) and a control group ($N = 43$). The experimental group engaged in a matching task with single sentences for each target vocabulary word. The target words were underlined in each sentence, while the control group did the same task without context, but with the Japanese meanings of target words. The overall result showed that there was no significant difference between contextual learning and decontextualized learning conditions.

Furthermore, Durrant and Schmitt (2010) showed that learning vocabulary in decontextualized sentences also had some effect on vocabulary learning. They investigated whether 84 non-native postgraduate ESL students could learn collocations in decontextualized sentences by associating target adjectives with nouns. They were divided into three groups to study target words: single-exposure, verbatim repetition,

and varied repetition groups. The single-exposure group saw 40 sentences only once containing 10 sentences with target nouns and their associative adjectives, 10 sentences with nouns without their adjectives, and 20 filler sentences. Each sentence appeared on a computer screen randomly and students read aloud the sentence for 7 minutes. The result showed that learners in the verbatim repetition and varied repetition group (Cohen's $d = 0.56$ and $d = 0.48$ respectively) outperformed those in the one-time presentation condition ($d = 0.25$). This suggests that learners could recall the best after being exposed twice with same sentences.

It should be mentioned however, that as the study of Durrant and Schmitt (2010) indicated, it should be mentioned that it is impossible to understand the exact meaning of PVs used in decontextualized sentences or situations because PVs can be used literally and metaphorically, which is same as collocations. This study thus employed sentences.

Moreover, it is important to draw learners' attention to target vocabulary by highlighting it when they read sentences. Without highlighting, learners may not recognize which words they need to focus on and learn. This approach is more likely to motivate learners to memorize the highlighted items in context (Boers & Lindstromberg, 2012). For instance, as demonstrated in Watanabe's study (1997), annotated vocabulary with marginal glosses was more easily recognized and led to

better incidental vocabulary learning than non-annotated words. Therefore, this approach should be applicable not only to single words but also to PVs. Highlighting may be more effective in helping learners learn PVs than simply presenting them as target vocabulary.

The Use of Visual Assists

In addition to the use of sentences, the use of visual aids, such as pictures and animations, has been identified as a useful tool in learning (Rieber, 1989; Wright, 1989). The use of pictures or animations can provide intrinsic motivation because learners can receive satisfying rewards from the enjoyable activity itself while doing a task (Ng & Ng, 2015; Schmidt et al., 1996). Paivio (1990) advocated Dual-coding theory, which hypothesizes theoretical benefits of the use of visual information from pedagogical perspectives. This theory assumes that the human cognitive system consists of two independent but interrelated functions. One is an image system that specializes in the processing of nonverbal information and is closely related to image processing. The other is a linguistic system that specializes in processing linguistic information. If memory is encoded using both code types and one of the codes is lost, the other code is still accessible. Thus, visual clues are highly conducive to long-term memory and this finding should be applicable to language learning. Although visual support should

theoretically assist the learning of PVs as well as single words, previous studies have never combined the use of visual information with learning PVs.

Several studies have investigated the effect of using pictorial information for incidental vocabulary learning while learners are listening or reading. For example, Bisson et al. (2015) investigated whether 53 ESL university students living in the U.K. could learn Welsh words through pictorial information while listening. The students were divided into two groups: an experimental ($N = 28$) and a control group ($N = 25$). There were three different learning phases, but the students in the control group were asked to start learning from the second phase. The first phase was an incidental learning, in which the experimental group was presented with a letter-search task in order to track their eye movements. They encountered all 78 target words six times, which contained six blocks for the 78 words. Each participant encountered 78 target Welsh words from one of three different lists, and 26 out of the 78 words were presented auditorily-only, 26 words were given auditorily with written participants' first language (L1) translations, and the rest of the 26 auditory words involved both L1 translations and pictures. In the second phase using explicit learning, both groups were given a translation recognition task and listened to a target word with its L1-written translation. The participants were told to judge whether the combination of a presented auditory word and its L1-translated meaning was correct or incorrect by pressing a button. After

each trial, a feedback was given. If they could not reach 80 percent accuracy, they were allowed up to three more trials. In the third phase, participants engaged in a recall and translation recognition test one week after the second learning phase. The recall test required students to listen to target Welsh words and type English translations of the auditory words while they were asked to complete the translation recognition test identically to the second phase without feedback. The results gained from a one-week later recall within the experimental group indicated that the accuracy rates on target auditory words with pictorial information and L1-translated meanings showed the highest retention (approximately 20 percent) than the words learned through L1-translated meanings-only (15 percent) and auditorily-only (13 percent). This seems to imply that pictorial information may contribute to learners' vocabulary learning even though the experiment concerned incidental learning through static pictures.

In another study conducted by Lin (2009), the effectiveness of dynamic and static annotations on motion verb learning was investigated. The study involved 70 Chinese eighth-grade students who were divided into three groups: 27 students with the use of dynamic animation, 21 with static graphics, and 22 with text-only. Twenty target words were selected from two children's stories, which the students had never encountered in their junior high school textbooks. The target words were highlighted in blue and underlined. Students in the dynamic animation and static graphics groups

could see an animation or graphics of a target word along with its Chinese translation by clicking it while reading the stories. The text-only group was presented with only the Chinese translation of a target word while reading the stories. Two different tests (production and recognition) were conducted before the experiment, immediately after the experiment, and two weeks after the immediate post-test. The production test required students to provide a correct letter of each target word with cues of its Chinese meaning and the first and last letter were given, whilst in the recognition test students were asked to select a a correct Chinese translation for the presented English word from four choices. The results showed improvements in performance for all three groups on the post-tests. Effect sizes were calculated to examine the magnitude of these improvements. In the immediate post-test, the static annotation group exhibited the highest effect size ($d = 2.03$), indicating a substantial improvement in vocabulary learning compared to the other groups. The animation group also demonstrated a moderate effect size ($d = 0.63$), while the text-only group showed a higher effect size ($d = 1.60$). However, it is worth noting that in the delayed post-tests, there were no significant differences between the groups ($d = 0.56$ for the static annotation group, $d = 0.48$ for the animation group, $d = 0.56$ for the text-only group). These findings highlight potential effectiveness of visual support and text presentation in enhancing vocabulary learning. They also imply that the learners' memory of the vocabulary items did not

endure until the delayed post-test, suggesting that the absence of retrieval practice may have contributed to this outcome.

The result of Lin's (2009) study revealed that inclusion of pictorial information could enhance incidental vocabulary learning during reading and listening tasks. However, it is important to note that the study focused on solely on incidental vocabulary learning while listening and reading and did not explore intentional learning with retrieval practices using pictorial information. This highlights a gap in the literature regarding the effects of pictorial information on intentional vocabulary learning, particularly with regards to PVs, and whether such effects can be sustained over the long-term.

Given the dynamic nature of PVs, it is possible that incorporating dynamic visual aids could provide additional benefits for students learning these expressions. As such, the present study sought to address this gap in the literature by utilizing short movies, or "PVgif" files, to illustrate the meaning of PVs. These dynamic visual aids are expected to enhance learning by reducing the cognitive burden associated with processing complex verbal expressions and promoting engagement and entertainment.

It is noteworthy that all of the previous studies cited have primarily focused on incidental vocabulary learning during reading or listening tasks. While the results suggest that pictorial information can contribute to vocabulary learning, no previous

research has examined long-term vocabulary learning, specifically related to PVs, using pictures or animations.

The Use of Tasks

According to Schmitt (2008, p. 339), any factor that increases the amount of exposure, use, time, or attention to vocabulary is likely to enhance learning outcomes. In this regard, additional activities and flashcard learning can increase students' exposure to the target vocabulary words, providing more opportunities for retrieval practice. This study employed drag-and-drop gap-filling (receptive) and translation (productive) tasks to promote intentional learning of PVs, using short example sentences that are easily accessible through students' smartphones, rather than reading or listening tasks.

Given the potential benefits of extracurricular learning, the present study utilized the *Moodle* system at a Japanese university to investigate whether extracurricular *Moodle* tasks can facilitate the learning of PVs.[4] The choice of *Moodle* as the platform for delivering the tasks was based on its popularity as a learning management system in

[4] *Moodle* is "a learning platform designed to provide educators, administrators and learners with a single robust, secure and integrated system to create personalized learning environments (Moodle, 2020)". It is known as one of the e-learning systems, which is the Learning Management System (LMS).

educational institutions and its user-friendly interface, which facilitates access to educational resources and communication among students and instructors.

E-learning

The present study aimed to investigate the effectiveness of utilizing tasks in *Moodle* for introducing example sentences and PVgifs to facilitate learning of PVs. In recent years, the use of the Internet has become a popular means of learning, providing learners with access to educational materials at any time and place. The COVID-19 pandemic has further highlighted the importance of e-learning and pedagogical internet tools in ensuring educational continuity.

E-learning has emerged as a popular learning method, enabling learners to access educational materials and resources via the Internet at any time and from anywhere. This growth has been facilitated by the proliferation of teaching and learning software, materials, and assessment tools available online. Large-scale e-learning systems such as WebCT and Blackboard have introduced innovative methods of learning such as online quizzes and collaborative learning through discussion boards. Recent studies have highlighted the significance of online learning in education, underscoring its potential to enhance learning outcomes and promote independent learning (e.g., Costa et al., 2012; Acar & Kayaoglu, 2020). The current global pandemic

has also emphasized the importance of e-learning as a reliable alternative to traditional face-to-face learning.

Moodle is one such e-learning platform that serves as a comprehensive Learning Management System (LMS), designed to create personalized learning environments for educators, administrators, and learners. *Moodle* offers a wide range of functionalities such as the ability to download materials, take quizzes, and submit reports. The platform has gained significant attraction among educators, particularly in higher education settings, where it is used to augment classroom-based learning or as a stand-alone educational program. *Moodle* is a valuable tool for educators, enabling them to assess student performance digitally, reduce paper-based assessment methods, and provide students with opportunities for independent learning through online quizzes and written assignments. Additionally, *Moodle* can be utilized to support vocabulary learning through deliberate exercises and tasks.

This study aims to investigate the implementation of the *Moodle* system in an actual vocabulary learning program at a university. Specifically, the study aims to explore the effectiveness of a voluntary, extracurricular learning activity through *Moodle* in students' learning of English PVs. While face-to-face learning has traditionally been the dominant mode of instruction, the rise of e-learning and distant learning has made it increasingly necessary to explore the potential of such methods in

supporting student learning. As such, this study seeks to contribute to the ongoing discourse on the role of e-learning in higher education.

3.2 Research Questions

The present study aimed to address the following research questions:

1. Does the use of extracurricular tasks combined with flashcard learning contribute to learners' ability to produce PVs in a production gap-filling test?
2. How much time do learners need to spend on the extracurricular tasks to improve their PV production ability in the production gap-filling test?
3. Are there particular types of extracurricular tasks that contribute more to learners' PV production ability than others?

3.3 Methodology

It should be noted that the methodology discussed in this chapter shares similarities with the methodology presented in Chapter 2. However, in order to avoid unnecessary repetition, the overlapping parts will not be restated here. Specifically, *Learning Program* remains the same, and a detailed report of it will not be provided in this section. For a more comprehensive understanding of the methodology, please refer to *Methodology* in Section 2.3, where detailed information can be found.

Participants

First, as to the number of participants that were needed in each group, an a-priori power analysis using G*Power (Faul et al., 2007) with parameters using a one-tailed hypothesis for matched pairs with an estimated effect size of $d = 0.7$, $\alpha = .05$, and power of $1\text{-}\beta = .95$ indicated that the group size needed to be at least 24 for each of the two groups.

The sample used in this study was identical to the one described in Experiment 1, as detailed in *Participants* of Section 2.3. The participants were initially divided into six different classes based on their performance in a TOEIC-like proficiency test that was administered at their university entrance and at the beginning of each semester. Students who did not take both the PV pre-test and post-test, as well as those who did not provide informed consent to participate in the study through *Google Forms*, were excluded from the analysis. The final number of participants included in the study was 125 students in the first semester and 103 students in the second semester.

Five classes participated in activities and one class did not receive any activities. The class that was not asked to engage in any activities was categorized as a control group (NO Moodle), and the students in the classes that had never engaged in them were also regarded as a control group (NO Moodle) because this *Moodle* learning was a voluntary extracurricular activity. On the other hand, the students who had participated

in the activities were in YES Moodle group.

The final number of the first-semester participants was 92 in YES Moodle group and 33 in NO Moodle group while the second-semester participants were 80 in YES Moodle group and 23 in NO Moodle group.

Moodle Activities

Moodle activities were set up to look at achievements between students who had extra time out of class and those who did not. Students in the five classes ($N = 105$) did extra activities whilst one class ($N = 20$) did not receive any activities for the first semester. On the other hand, in the second semester, students in the five classes ($N = 90$) received the activities while those in one class ($N = 13$) were not given the activities. The activities were voluntary, so the students who did not engage in them were classified into the control NO Moodle group. Other students were given opportunities to access to different activities weekly, and fifteen minutes at the most were provided for one trial. The weekly activity was accessible for one week after they learned new five PVs, and the students were allowed to practice the five words as many times as they wanted until the activity was shut down after one week. After that, they were able to review their previous trials. However, the study was not able to control the amount of time each participant spent on the activities due to the very nature of the voluntary activities.

Tasks on Moodle Learning

Students in certain classes were informed that participation in *Moodle* activities was an essential aspect of their course, while students in other classes were given the option to participate. The *Moodle* activities were made available to students during class time, subject to the discretion of their respective teachers. Five out of six classes were randomly assigned to different types of PV learning tasks: translation, gap-filling, and gap-filling with short animation PVgif. The three types of PV learning tasks used on *Moodle* were as follows:

1. Translation task: See a sentence and type in the correct PV by translating the part of target PV from English to Japanese and Japanese to English--both types.

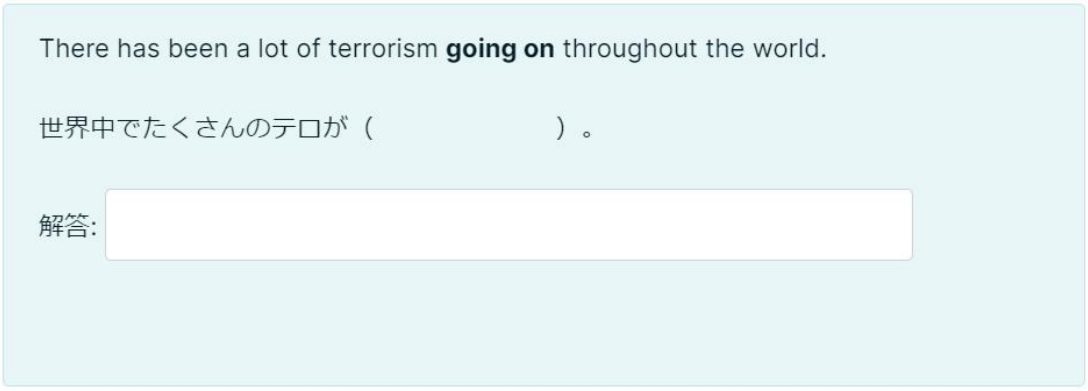
2. Gap-filling task: See a sentence and choose the appropriate verb and adverbial particle respectively in a sentence from a list including five verbs and thirteen adverbial particles.

3. Gap-filling with PVgif task: See a PVgif with a sentence and select the proper verb and adverbial particle for the sentence from a list including five verbs and thirteen adverbial particles.

The first type of *Moodle* activity for the Translation group needed to translate the Japanese part of the target PV into English, and vice versa.

Figure 5. *Example of Japanese-to-English Translation Task on Moodle*

Figure 6. *Example of English-to-Japanese Translation Task on Moodle*

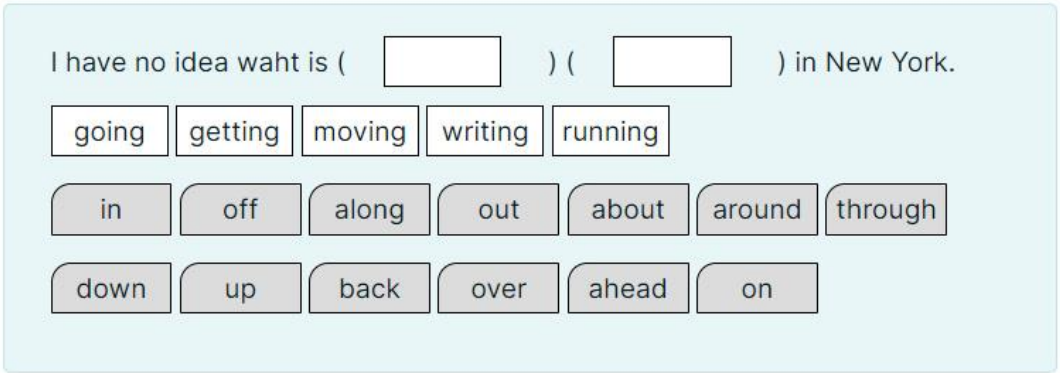


For example, in Figure 5, students were initially presented with both an English sentence and its Japanese translation, but only the English sentence contained blanks representing a target PV that students were required to translate. The answer should be in the present progressive tense, “going on.” On the other hand, in Figure 6, only the

Japanese sentence had a blank, and students had to translate the target PV in the English sentence into the present progressive tense in Japanese, “起こっている.”

In the second *Moodle* activity, referred to as the Gap-filling activity, students were presented with sentences and were required to choose the appropriate verb and adverbial particle from a given list. The verbs and adverbial particles were selected from a pool of options, all of which were included in the PHaVE List (Garnier & Schmitt, 2014). The students had to select the correct verb and adverbial particle carefully from five verbs and thirteen particles provided in the list. To complete the activity, they would then drag and drop the selected items into the appropriate spaces in the sentence (Figure 7).

Figure 7. *Example of Gap-Filling Task with Drag-and-Drop from List on Moodle*



The third *Moodle* activity was watching PVgifs; short, animated clips that illustrated the meaning of PVs using verbs and adverbial particles. These clips were

distributed as compressed GIF files, which reduced the bandwidth requirements for students accessing the *Moodle* activity. The PVgif task was designed as a gap-filling exercise that required students to select the correct verb and adverbial particle from a list by dragging and dropping on the *Moodle* screen (Figure 8).

Figure 8. *Example of Gap-Filling Task with a PVgif on Moodle*

Over the course of one academic year, I created a total of 120 PVgifs in collaboration with native teachers of English and Japanese students at university. The PVgifs primarily depicted dynamic verbs, showing actions and movements associated with PVs. This collaborative process involved working closely with the teachers to identify the dynamic meaning of each PV, and then working with the teachers and students to create PVgifs that were easily understood by the participants. The students involved in creating PVgifs were not included in the study.

All *Moodle* activities consisted of 10 questions every week, and the students were required to address each of the five PVs twice. The same sentences were used in all three conditions. Students could access the activity every week.

Phrasal Verbs

The selection of PVs follows the same approach as described in Chapter 2. Specifically, the study chose the 120 most frequent PVs from the PHaVE List for one academic year. Additionally, 840 academic vocabulary words were chosen from the Academic Vocabulary List (AVL) for the university's vocabulary learning program. For further details regarding the selection process, please refer to *Phrasal Verbs* in Section 2.3. Note, incidentally, this study focused on learning PVs and did not analyze learners' outcomes on academic vocabulary.

Sentences

All the sentences used in the *Moodle* activity and two tests were obtained from the *Corpus of Contemporary American English* (COCA) (Davies, 2008-). To ensure the authenticity and relevance of the sentences, each one was designed through a collaborative effort with a native speaker who has expertise in the field of linguistics. Authentic sentences is preferable to artificially created ones, as they better reflect the

nuances of the word in context. However, to ensure that the sentences were understandable outside of their original context, slight modifications were made to some of the sentences. These modifications were carried out carefully to avoid altering the meaning or structure of the sentences, while enhancing their clarity for learners. PHaVE Appendix A provides a list of the sentences used in the tests, while PHaVE Appendix C lists the sentences used in the activities. By using authentic sentences that were modified for clarity and comprehensibility, learners were able to engage with and develop a deeper understanding of the complexities of English PV usage.

Furthermore, to draw learners' attention to the target PVs, they were highlighted and presented within example sentences in context through *Moodle* activities, but not in tests.

Tests

This section will briefly describe the test methodology as it remains the same as Experiment 1. Please refer to *Tests* in Section 2.3 for more details.

The test methodology involved administering a pre-test and a post-test to all participants using *Google Forms*. It consisted of a fill-in-the-blank task with 50 PVs selected from the targeted 60 PVs for each semester. To aid participants in their responses, the pre-test provided the first two letters of the PVs and a Japanese

translation as hints. The Japanese translation was not highlighted in any way. This was done to get rid of the contextual variations that could lead to non-target PVs or potentially correct answers, such as *get in*, *go in*, and *come in*. The number of words and verb tense in Japanese were also displayed beneath the English sentence to ensure desired answers.

3.4 Results

The data were subjected to a Shapiro-Wilk normality test to determine if they followed a normal distribution prior to conducting any statistical tests. The results indicated normal distribution in two conditions and three tasks. A significance level of .05 was adopted to test for statistical significance. A one-way ANOVA was carried out to examine the differences among the three tasks, followed by post-hoc tests to determine specific group differences.[5] Additionally, effect sizes were calculated to determine the magnitude of the differences observed.

[5] One-Way Analysis of Variance (ANOVA): "this is used when you want to test whether the scores of three or more groups or levels differ statistically" (Larson-Hall, 2015, p. 180).

3.4.1 1st Semester

Moodle Activities

The students who participated in extra *Moodle* activities with the PVs showed any advantage over the students who did not do this. Descriptive statistics on students' gain scores during the first semester are presented in Table 4.

Table 4. *1st Semester YES Moodle vs. NO Moodle Group Gain Scores*

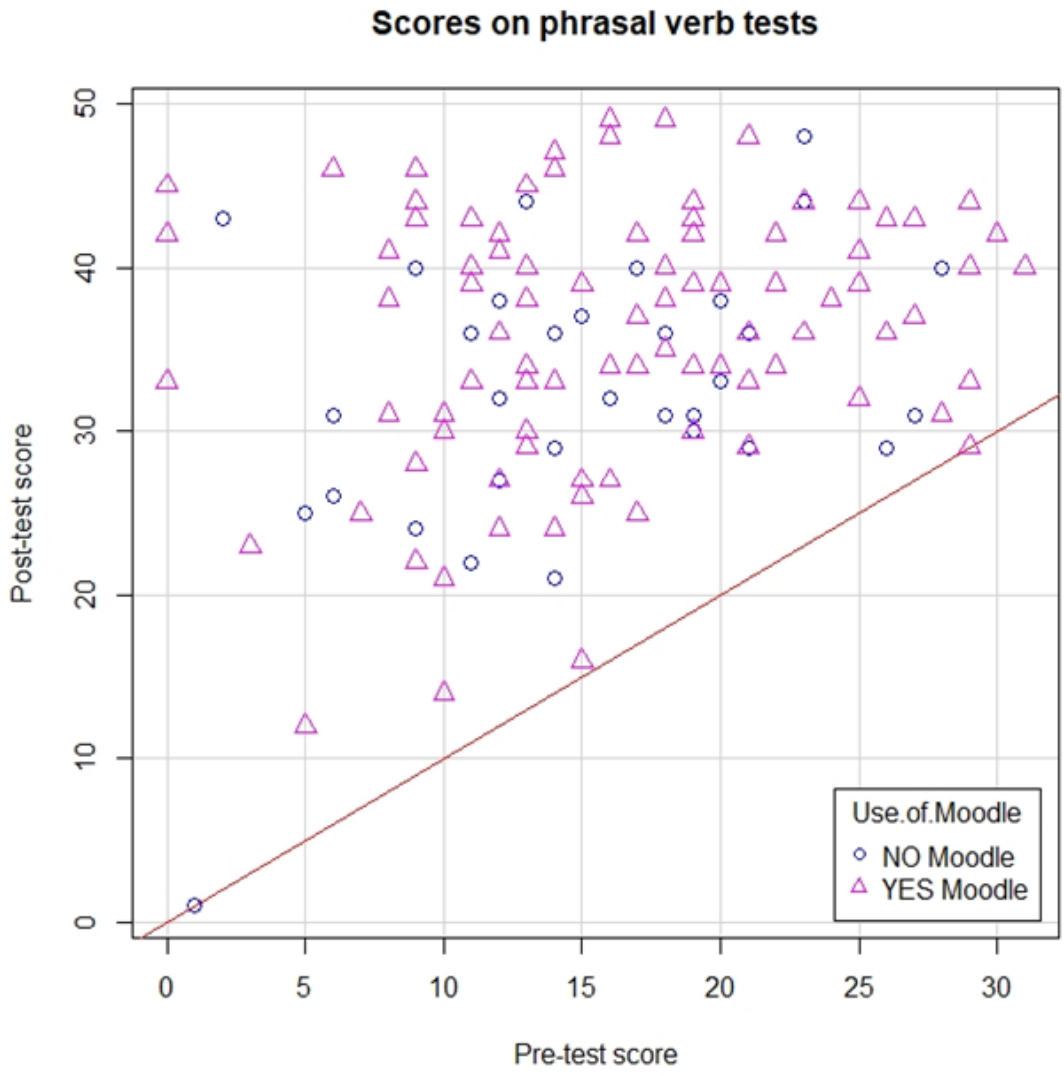
Moodle groups	N	Mean	SD
YES Moodle	92	19.79	9.49
NO Moodle	33	17.85	8.93

The group who participated in *Moodle* had a slightly higher mean score than the group who did not. However, an independent samples *t*-test with the gain scores revealed no statistically significant difference between the YES and NO Moodle groups ($t(59.747) = -1.06$, $p > .05$ (= .30)), with a small Cohen's *d* effect size of $d = 0.21$. Furthermore, the 95% confidence interval (CI) also showed no effect between the two conditions (95% CI: [-5.63, 1.74]).

A scatterplot in Figure 9 shows the correlation between pre-test and post-test scores for all 125 students. Most students exhibited progress regardless of *Moodle* activities, falling above a 45-degree diagonal line. However, a negligible correlation

was found ($r = 0.27$).

Figure 9. *Correlation of Gain Scores on Phrasal Verb Gap-Filling Pre-Test and Post-Test in 1st Semester*



The lack of a strong correlation between the use of *Moodle* activities and student performance may be attributed to the nature of the activities, which were designed to take no more than 15 minutes per week and were often completed in less time. This

suggests that the *Moodle* exercises may not have been sufficient to produce a significant change in student performance. Furthermore, it should be noted that even the students who participated in *Moodle* activities for only a brief period during the semester were included in the "YES Moodle" condition, which may have influenced the overall results.

Upon analyzing Figure 9, which distinguishes the students who used *Moodle* from those who did not with different markers, there was no observable disparity in the scatter between the two groups. Given that the *Moodle* activities were voluntary, many students participated with a lesser degree of involvement. Consequently, the present study undertook an investigation into the correlation and regression between the frequency of students' completion of the weekly activities, their gain scores, and the total duration spent on the tasks.

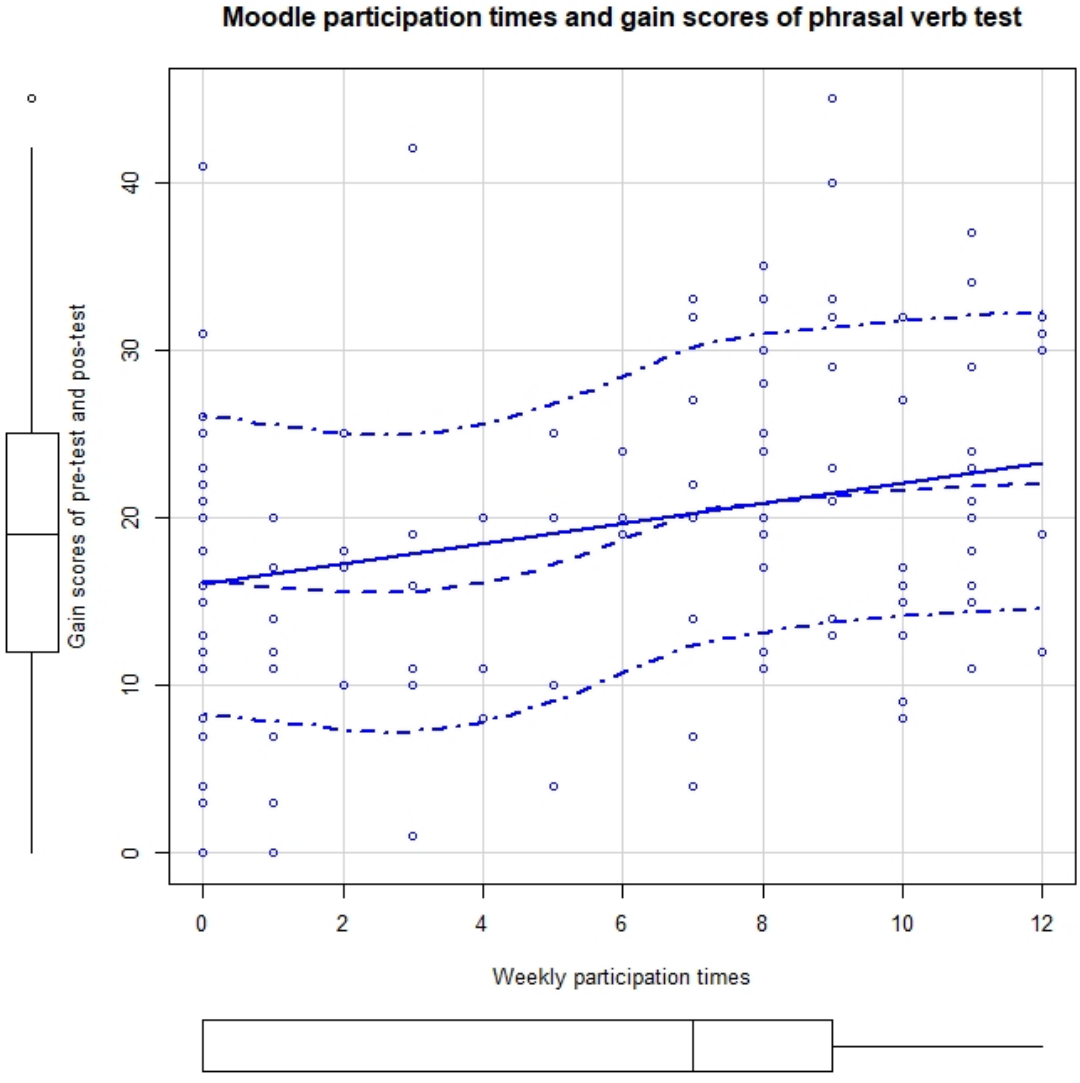
Participation Times on Tasks

The scatterplot illustrated in Figure 10 presents the relationship between the number of weeks that participants engaged in *Moodle* activities and their gain scores.

A Pearson's correlation coefficient showed a low correlation between the number of weeks that students utilized *Moodle* and their gain scores on the PV tests, with a coefficient of 0.28. Furthermore, a relatively weak correlation was found

between the total minutes that students engaged in activities throughout the semester and their gain scores, with a coefficient of 0.22.[6]

Figure 10. *Scatterplot of Weekly Attempts on Moodle Activities and Gain Scores in 1st Semester*



[6] Note that although the correlation coefficients are similar, this is a different correlation than the one above between pre-test and post-test score.

A line slope in Figure 10 is positive and there seems to be a positive linear relationship between the students' gain scores of the pre-test and post-test and their number of weekly trials. A loess line was placed on the plot to determine a cut-off point for additional learning effects from extra practice. However, it was challenging to pinpoint a distinct threshold, making any analysis of whether extra practice benefited students based on whether they participated.

Task Types

In order to see whether the types of activities done on *Moodle* produced different results, Table 5 presents descriptive statistics on the types of activities.

Table 5. *1st Semester Gain Scores by Task Type*

Task types	N	Mean	SD
Translation	33	20.58	9.90
Gap-filling	37	18.86	9.06
Gap-filling + PVgif	22	20.18	9.87
NO Moodle	33	17.85	8.93

The Translation task produced the highest gain score (M = 20.58, SD = 9.90), followed by the Gap-filling + PVgif task (M = 20.18, SD = 9.87), and the Gap-filling task produced

the lowest gain score (M = 18.86, SD = 9.06).

A one-way ANOVA was conducted to consider the factor of the three task types. It revealed no significant difference among the three types of tasks since the p-value was more than 0.5 ($F(2, 89) = 0.30$, $p = .74$). The pre-test and post-test were both gap-filling tests, and the results revealed no statistically significant difference between the three tasks at the $p = .05$ level. Therefore, there is no evidence to suggest that one task resulted in better learning outcomes than the other tasks.

Table 6 displays the results of Tukey's post-hoc comparisons between the three tasks. It revealed that there was no statistical difference between the tasks at the $p = .05$ level. The Cohen's d effect sizes show a $d = 0.18$ for Translation vs. Gap-filling tasks, $d = 0.04$ for Translation vs. Gap-filling + PVgif tasks, and $d = 0.14$ for Gap-filling vs. Gap-filling + PVgif tasks. The effect sizes of the three tasks were quite small. The three tasks could not be differentiated with respect to their additional impact on learning PVs and all the tasks had some effect on students' learning of PVs regardless of the use of PVgifs.

To avoid the risk of making false discoveries when testing multiple hypotheses, a statistical method known as False Discovery Rate (FDR) correction was applied. The FDR values obtained for all three pairwise comparisons were 0.99, which indicates a high probability of finding false positives among the significant results.

Table 6. *Multiple Comparisons of Task Types in 1st Semester*

Task types	*t* value	Pr(>|t|)	95% CI		FDR	Cohen's *d*
			Lower	Upper		
Translation : Gap-filling	0.75	0.74	-3.75	7.17	0.99	0.04
Translation : Gap-filling + PVgif	0.15	0.99	-5.88	6.67	0.99	0.18
Gap-filling : Gap-filling + PVgif	0.51	0.87	-4.82	7.45	0.99	0.14

Additionally, the regression analysis was performed to explore the relationship between post-test scores and multiple predictor variables, including pre-test scores, task types, and total time on task. This analysis allows us to understand the collective impact of these variables on learning outcomes and provides a more comprehensive understanding of the factors influencing vocabulary acquisition. By incorporating the one-way ANOVA and *t*-test alongside the regression analysis, a more nuanced perspective on the relationships and effects observed in this study was attained.

Regression Analysis of Predictors on Post-test Scores

The presented regression analysis aims to identify the relationship between post-test scores and pre-test scores, task types, and time on tasks for the first semester. Table 7 provides the unstandardized coefficients (*B*), standard errors of the coefficients (*SE B*),

standardized coefficients (β), dominance weights, and 95% CI for each predictor variable. The results show that pre-test scores ($B = 0.31$, $\beta = 0.27$, $p < .01$, 95% CI: [0.12, 0.51]) and time on task ($B = 3.92$, $\beta = 0.30$, $p < .01$, 95% CI: [1.23, 6.61]) are significant predictors of post-test score, whereas task types ($B = -0.41$, $\beta = -0.05$, $p = .62$, 95% CI: [-2.00, 1.19]) is not a significant predictor.

Table 7. *Regression Results for Predictors of Post-Test Score in the 1st Semester*

Variable	B	SE B	β	Dominance Weight (%)	95% CI	Rank	Boruta Predictor Importance
Pre-test score	0.31	0.10	0.27	.071 (48.63)	[0.12, 0.51]	1	Rejected
Time on task	3.92	1.36	0.30	.065 (44.52)	[1.23, 6.61]	2	Confirmed
Task types	-0.41	0.80	-0.05	.010 (6.85)	[-2.00, 1.19]	3	Rejected
Total				.146 (100%)			

Note. Criterion variable is post-test scores, $N = 125$, Multiple $R^2 = .146$, Adjusted $R^2 = .125$, F-statistic = 6.88, p-value < .001, Residual standard error = 7.761 on 121 degrees of freedom.

The results were submitted to Relative Weight Analysis on langtest.jp (Mizumoto, 2015). Boruta feature selection analysis confirmed time on tasks as an important predictor of post-test scores. The model explains 14.6% of the variance in post-test scores ($R^2 = .146$) and has a significant overall fit ($F(3, 121) = 6.88$, $p < .001$). The adjusted R-squared is .125, indicating that 12.5% of the variance in post-test scores

can be explained by the predictors after adjusting for the number of predictors in the model. The residual standard error is 7.761 on 121 degrees of freedom, suggesting a good fit of the model to the data.

These findings suggest that time on tasks is an important factor that contributes to student performance in the first semester, while pre-test scores and task types have limited influence on post-test scores. Pre-test scores serve as a nuisance factor that accounts for differences in starting points, but the regression analysis effectively isolates or factors this out for the dependent variable. The lack of significance for task types indicates that it did not significantly affect the results, or that there was insufficient power to detect a significant effect. The effect size of time on tasks was found to be significant in the regression analysis and can provide insights into the relative importance of this factor.

In conclusion, the results of the regression analysis indicate that pre-test scores and time on tasks significantly impact student performance in the first semester. The findings suggest that these factors play a meaningful role in determining post-test scores. However, the influence of task types on post-test scores was found to be relatively limited. Although time on tasks had a discernible effect, it was not substantial enough to identify a specific type of activity that significantly enhanced students' learning of PVs.

3.4.2 2nd Semester

To highlight the key findings and improve clarity, the presentation has been streamlined to prioritize the regression analysis. The focus has been placed on the most relevant information regarding the relationship between predictor variables and post-test scores, as the results for the second semester align with those of the first semester that have already been analyzed and documented in detail. This simplification aims to enhance understanding and avoid unnecessary complexity, ensuring that the key findings are effectively communicated.

Regression Analysis of Predictors on Post-Test Scores

Table 8 displays the results of the regression analysis, examining the relationship between predictor variables and post-test scores in the second semester. The findings indicate that pre-test scores ($B = 0.54$, $\beta = 0.40$, $p < .001$) and time on tasks ($B = 4.92$, $\beta = 0.34$, $p < .001$) were significant predictors of post-test scores, while task types ($B = -0.06$, $\beta = -0.01$, $p = .94$) was not a significant predictor. The adjusted R-squared was 0.259, indicating that 25.9% of the variance in post-test scores could be explained by the predictors. The F-statistic was significant ($F(3, 99) = 12.35$, $p < .001$), indicating that the regression model as a whole was significant.

Table 8. *Regression Results for Predictors of Post-Test Score in the 2nd semester*

Variable	B	SE B	β	Dominance Weight (%)	95% CI	Rank	Boruta Predictor Importance
Pre-test score	0.54	0.11	0.40	0.16 (59.26)	[0.31, 0.77]	1	Confirmed
Time on task	4.92	1.35	0.34	0.10 (37.04)	[2.24, 7.59]	2	Confirmed
Task types	-0.06	0.70	-0.01	0.01 (3.70)	[-1.44, 1.33]	3	Tentative
Total				0.27 (100%)			

Note. Criterion variable is post-test scores, $N = 103$, Multiple $R^2 = .272$, Adjusted R-squared = .25, F-statistic = 12.35, p-value < .001, Residual standard error = 7.659 on 99 degrees of freedom.

These results highlight the importance of pre-test scores and time spent on tasks in predicting student performance in the second semester. Specifically, higher pre-test scores and increased time dedicated to *Moodle* activities are associated with higher post-test scores, underscoring the effectiveness of additional activities in enhancing student learning outcomes.

3.5 Conclusion

The aim of this study was to investigate a potential contribution of additional activities to learners' PV learning. The results revealed that learners who participated in the extra activities achieved slightly higher gains compared to those who did not, although the difference was small. Surprisingly, there was no significant effect observed for the type

of task, even when visual assists were included.

Furthermore, the study uncovered a weak correlation between the number of learners' weekly participation in *Moodle* and their gain scores on the PV tests, without a clear threshold indicating optimal participation. However, the regression analysis indicated that dedicating more time to additional activities was associated with higher post-test scores. These findings suggest that engaging in extra activities on *Moodle* can have a positive impact on students' learning outcomes.

3.6 Discussion

The findings of this study suggest that the additional activities had a limited impact on PV learning. The lack of significant differences between task types indicates that visual assists did not significantly contribute to improved learning outcomes. These findings are in disagreement with previous studies that suggested visual assists could be helpful tools for vocabulary learning (Bisson et al., 2015; Lin, 2009). However, the regression analysis conducted on the results of this experiment revealed that the total time spent on tasks emerged as an important factor influencing student performance. This suggests that learners' engagement with tasks and the amount of time dedicated to them played a crucial role in learning PVs. The results imply that the quality of learning and engagement tends to be influenced by the total time spent on a task, rather than the

number of participations. In contrast, the findings of Hill and Laufer (2003) indicated that the time required to complete tasks (form-oriented and meaning-oriented tasks) did not have a significant difference although their focus was on incidental vocabulary learning. Instead, they discovered that actively engaging with words, exploring their forms, meanings, and relationships, led to higher levels of word-related activity that facilitate vocabulary acquisition and retention. The absence of consensus in previous research regarding the effect of time on tasks on vocabulary learning emphasizes the complex nature of this relationship and the need for further investigation.

Based on the findings, it can be concluded that additional activities for PVs may have a limited impact, but their role should not be disregarded in comparison to solely memorizing word definitions. Simply knowing the definition of a word is insufficient for a comprehensive understanding of its meaning. Effective vocabulary learning requires active engagement, involving exploration of how words are used in different contexts and their relationships with other words (Stahl & Kapinus, 2001). Active engagement helps learners develop a deeper understanding and the ability to use words appropriately in various scenarios. This active engagement can include activities like extensive reading, discussing word usage, analyzing texts, and practicing using words in output. Laufer (2003) supports the effectiveness of activities such as sentence completion, writing words in sentences, and incorporating them into compositions for

vocabulary learning, surpassing the passive encounter of words through reading. These activities proved more effective in acquiring and retaining vocabulary. However, it is crucial to recognize that a single session of form-focused vocabulary learning is not enough for comprehensive learning and long-term retention (Webb et al., 2020). While such a session may yield significant initial gains in understanding the connection between word form and meaning, it should be viewed as the initial stage of the word learning process. Continued practice and engagement are necessary to ensure comprehensive understanding and long-term retention.

Although no specific research has been conducted on the use of additional tasks in conjunction with flashcard learning to enhance vocabulary learning, including PVs, ideally, educators should incorporate a variety of activities that promote active engagement and meaningful interactions with words. This view is supported by studies conducted by Webb (2005), Nassaji and Tian (2010), and Teng (2017), which highlight the benefits of incorporating output tasks and meaningful language use in vocabulary learning. Nassaji and Tian (2010) and Teng (2017) emphasize the importance of active engagement through collaborative work with other students in PV learning. These experiments demonstrate that such activities provide valuable opportunities for learners to practice and internalize new words, thereby enhancing their vocabulary acquisition. Therefore, I recommend adopting a multifaceted approach that includes active activities,

promoting a comprehensive and engaging vocabulary learning experience for learners.

Further research is needed to investigate the relationship between additional tasks and long-term retention of PVs. Additionally, future studies could explore consistent and continuous learning with required extra activities to determine their impact on PV acquisition. By addressing these gaps, a more comprehensive understanding of the effectiveness of additional activities in PV learning can be obtained.

Several limitations should be acknowledged in this experiment for future research. Firstly, regarding the use of *Moodle*, most students only engaged in the activities once or twice a week, or for less than an hour throughout the academic semester. I should note that this limited engagement could be attributed to the activities having only ten questions and a repetitive structure. It remains uncertain whether better results could have been achieved if students had devoted themselves to more earnest study of the extra materials. Additionally, measuring students' study time based on their *Moodle* window being open may not accurately reflect their actual time spent on the activities.

In terms of the tests, both the pre-test and post-test were accessible over several days, allowing the students to access the target PVs and their meanings from class resources while taking the test. This may have influenced the results. Although this

situation was unavoidable due to the COVID-19 pandemic-related university closures, I should mention that conducting the tests under the supervision of a teacher or implementing time limits would have been ideal for ensuring secure test administration. Furthermore, to improve the control over the recall output in a test, it could be advantageous to use blanks to indicate the number of letters in each word or PV and provide a few hint letters alongside the first one or two letters of the verb. Such a testing format could help learners concentrate on recalling the spelling of the word, rather than recognizing it from a list of options. This approach may also bring about a more precise evaluation of their vocabulary knowledge. Moreover, the format of the PV tests and gap-filling task conditions were quite similar, which may have influenced the absence of an effect observed in the Gap-filling with PVgif task condition. This similarity could have led students to approach both tasks in a similar manner, regardless of whether visual assists were included or not.

Furthermore, in terms of groups, the experimental design did not have a control over a potential influence of the teacher on the results. In this study, one class served as the control group (referred to as the “NO Moodle” group) as they did not receive any activities. Additionally, students from other classes who did not engage in activities were also included in the control group. However, the design did not account for variations in teaching methods or approaches among the teachers. Therefore, a potential

impact of the teacher's effect on the results could not be controlled or accounted for in the study. It is worth considering this limitation when interpreting the findings and generalizing the results to other contexts.

Lastly, I should acknowledge that the type of tasks varied between two academic semesters and classes. As a result, this study was unable to investigate the relationship between task types and students' retention of PVs. Future research might be necessary to look at consistent and continuous learning with extra activities.

In conclusion, this study suggests that additional activities alongside flashcard learning may have a positive impact on learning PVs, particularly in terms of time spent on the activities. However, further research is needed to fully understand their effectiveness and develop engaging activities that promote active engagement for optimal learning outcomes.

CHAPTER 4.

Experiment 3: A Narrative Task Approach for Learning Phrasal Verbs

One of the ultimate goals of language learners is to become native-like users of the target language. However, achieving this level of proficiency can be a daunting task, particularly when it comes to mastering formulaic sequences (FSs) such as idioms, collocations, and phrasal verbs (PVs). PVs, in particular, pose a significant challenge to learners, as they are frequently used in spoken English and can have multiple meanings depending on the context.

During my pursuit of English proficiency, I found myself constantly struggling to understand and use PVs, even in relatively informal communication such as email exchanges with native speakers of English. I have encountered confusing PVs that they use very naturally, such as "let's *hash* this *out*" and "the information will be *rolled out* next month." I often wondered how they come up with so many PVs. LX learners are much less likely to use PVs than native speakers (for the term 'LX,' see Footnote 1). In speaking situations, especially on the fly, it is quite difficult for learners to understand the meanings of PVs. This experience prompted me to explore ways to help learners like myself move beyond mere acquisition to actively using PVs in speaking.

In this chapter, I will describe my study on the effectiveness of using narrative tasks that involve telling a story as an approach to learning PVs. Specifically, the study

aimed to investigate whether adding a speaking component to PV memorization using computer flashcards could help learners use more PVs in a speaking task, and to recall and recognize PVs measured by a gap-filling test. The findings of this study will shed light on the potential of incorporating speaking tasks into PV instruction to promote more authentic and effective language learning experiences.

4.1 Literature Review

Learning with Flashcard Learning Program Anki

To study phrasal verbs (PVs) for use in speaking, the flashcard learning system *Anki* was incorporated in this study. Previous research led me to assume that this type of approach would result in gains for students who consistently studied their words. In particular, Chapter 2 showed that students made great progress in acquiring PVs over one academic semester (from 30 % to 70% of PVs in the first semester and from 50% to 64% in the second semester) and successfully retained about 50% of the total number of PVs that the students were tested on each semester even one year after they had finished learning the PVs. These results indicate that spaced retrieval practices are crucially effective to vocabulary learning.

One issue that needs to be addressed when doing a study with retrieval of vocabulary is how many times participants should see any individual word. Although the exact number of repetitions needed for learning new vocabulary is still controversial, retrieving meanings of vocabulary over time logically appears to be a superior way of learning vocabulary. The *Anki* program uses a spaced repetition algorithm as a default that shows a card again in 1 minute, then 10 minutes, then 4 days, 10 days, 25 days, and 75 days, if the user gets the word correct every time. Thus, in this 9-week study, even a word at the end of the study period was seen at least three times

before the immediate post-test. Obviously, if the learners did not know the word, they would see it more times. The results of this study would be thus based on the premise of spaced repetition flashcards that function in *Anki* by helping LX users gain vocabulary in general (see Narrative Appendix A for the list of the words included in this study). This fact was assumed, not tested, so the real question for the study was whether output activities, combined with memorization practices, could enhance students' ability to use PVs in their spoken language.

Collaborative Learning on Speaking

Considering the role of output in the vocabulary acquisition process, it has been hypothesized that output can serve as a crucial step for enabling learners to identify gaps between what they *want to say* and what they *can* say. In this regard, negotiated output, which involves interactions among learners, has been shown to be particularly effective for incidental vocabulary acquisition (Ellis & He, 1999). By engaging in learner-to-learner interactions, learners are prompted to actively notice and address the gaps in their vocabulary knowledge, leading to enhanced acquisition outcomes.

A few studies have investigated whether pair work on a task is more effective than individual work by using PVs. The focus in these studies was on how task type and interaction affected learning outcomes. One of the studies that incorporated learning of

PVs was Nassaji and Tian (2010). Their study suggests that spoken output and interaction can be useful for acquiring PVs. They investigated whether collaborative or individual work of 26 low-intermediate ESL students on two written tasks over 2 weeks would affect the learning of 16 PVs. One is a cloze task where students needed to listen to a passage and reconstruct it, while the other is an editing task where students had to find and correct errors in a written text based on a dialogue they had just heard. All the participants completed the different two tasks (cloze and editing tasks) with both collaborative and individual works. Each task contained 4 PVs from the 16 target PVs and limited to 8 minutes (4 PVs x 4 tasks). Results were measured by looking at the accuracy of the PVs for each task and scores on a vocabulary knowledge scale before and after the two days of class in which the tasks were conducted. The students worked on these tasks for two consecutive weeks. The results showed that gains and effect sizes for collaborative works (M = 3.90 and Cohen's d = 1.60 for the editing task and M = 2.70 and d = 0.96 for the cloze task) were larger than individual works (M = 3.05 and d = 1.32 for the editing task and M = 1.90 and d = 0.72 for the cloze task).

Another relevant study on the acquisition of PVs was conducted by Teng (2017). The study aimed to examine the effects of different tasks on PV knowledge improvement over a three-week period. The study utilized short editing and cloze tasks, as well as a writing task, and compared individual, pair, and group work conditions.

This design resulted in a total of nine combinations of task-condition conditions, with 24 participants in each, yielding a total of 72 participants. Each participant completed all three tasks, one in each type of condition. The participants were required to learn 18 PVs, suggesting that each task likely included six different PVs. To assess participants' PV knowledge, scores were awarded based on their ability to provide L1 translations or synonyms for the verbs and to use the verbs appropriately in sentences, as measured by a vocabulary knowledge scale. The data were analyzed using a series of repeated measures ANOVA (RM-ANOVA), considering the factors of task, condition, and time (pre-test vs. post-test) for each task separately. The results indicated a significant statistical interaction between time and condition in all three tasks. Post-test scores were higher, and post-hoc comparisons revealed that the group condition scores surpassed the pair condition scores, which, in turn, were higher than the individual scores. Thus, overall, participants demonstrated greater PV knowledge gains through collaborative work compared to individual work. Another RM-ANOVA, which did not differentiate between conditions, examined the interaction between task and time. The analysis revealed a significant statistical interaction, with higher post-test scores. Post-hoc comparisons indicated that the writing task yielded higher learning scores compared to the editing task, which, in turn, outperformed the cloze task. It is important to note,

however, that these findings did not yield significant vocabulary acquisition within the three-week learning period.

Both of these studies, the Nassaji and Tian (2010) study and the Teng (2017) study, thus imply that working with other students together would enhance their knowledge of PVs. However, these studies were relatively short-term, spanning only two or three weeks. The current study thus decided to employ pairs of learners speaking together and giving each other negotiated feedback. In addition, the studies suggest that short-term interventions can improve knowledge of PVs. It is unclear, however, whether such short-term improvements would result in long-term gains in real vocabulary. Thus, in the present study, learners were told to study PVs every week used a spaced repetition system along with a regular speaking component in order to investigate how a memorization component could affect the acquisition and spontaneous production of PVs. This study was a longitudinal one, examining how addition of a regular speaking component on top of a memorization component improves retention and spontaneous production of PVs.

Taking into consideration a speaking task that could incorporate PVs, this study used a picture description activity. Although one study by Derwing et al. (2006) used a picture story, it was for measuring English pronunciation and fluency of Mandarin and

Slavic speakers. It might be rare for studies to use narrative tasks as a method of learning within the context of L2 instruction.

Storytelling is a social activity that frequently happens in our daily lives and plays a crucial role in constructing our conversation (Wong & Waring, 2010). This would also be applicable to language learning--the ability to make and tell a story can lead to interaction and improve communication skills. However, to my knowledge, no research has focused on whether vocabulary retention could be improved by asking participants to both memorize and use their memorized vocabulary in speaking tasks.

In the present study I wanted LX users to be able to use PVs in speaking as well, so this study encouraged LX users to produce PVs in weekly speaking tasks. Using a speaking task also invokes Swain's (1985) Output Hypothesis, which has been shown in numerous studies to correctly predict that producing output results in better outcomes than situations where output does not need to be produced (Boers et al., 2006; Mackey, 2006; McDonough & Mackey, 2000).

Involvement Load Hypothesis

The results of the studies conducted by Nassaji and Tian (2010) and Teng (2017) provide empirical evidence supporting the effectiveness of incorporating PVs in short-term interventions using various task types, such as collaborative and individual tasks.

These findings support the Involvement Load Hypothesis proposed by Laufer and Hulstijn (2001), which suggests that tasks involving more indices of involvement, such as search, need, and evaluation, can increase the likelihood of vocabulary retention. In this context, the use of PVs in speaking tasks can provide learners with an opportunity to practice and use the targeted vocabulary in a meaningful context, thereby enhancing their involvement and retention of these items over time. Several studies have found that although not perfect, this hypothesis seems to be a good way to predict how well different types of tasks will perform in helping language learners to remember vocabulary (Keating, 2008; Kim, 2008; Laufer, 2003).

After evaluating Nassaji and Tian's (2010) study on the involvement load of a cloze and editing tasks, I concluded that the cloze task has high++ need for specific words, no search, and moderate+ evaluation. In comparison, the editing task has a high++ need, no search, and high++ evaluation. The study found that the effect sizes for the editing task were larger and statistically different from those of the cloze task. The study suggests that short-term interventions can enhance knowledge of FSs when using tasks that demand higher levels of involvement.

Teng (2017) conducted a different study using short editing and cloze tasks, as well as a writing task to study improvements in PV knowledge. When the involvement load hypothesis was applied, and the editing task was expected to require more

involvement than the cloze task. The writing task, which asked participants to use PVs to summarize a passage they had heard, would be similar to the editing task in involvement (high need ++, no search, high evaluation ++). The study found that the writing task had higher learning scores than the editing task, which was higher than the cloze task.

Both of these studies thus imply that tasks with higher involvement load should lead to better retention of FSs. In a speaking task, the speaker need to evaluate whether the word fits the context. Laufer and Hulstijn (2001) also note that negotiated interaction implies a need for the search of meaning, which fits with Teng's (2017) better results with collaboration.

However, it should be noted that there is a potential problem in estimating the involvement load levels of tasks, as different researchers may use different criteria. In the present study, the involvement load levels of the tasks used in previous studies could not be determined unambiguously. For example, when using the involvement load hypothesis suggested by Laufer and Hulstijn (2001), the involvement load levels of the editing task used by Nassaji and Tian (2010) and Teng (2007) could be estimated as the same as the writing task used by Teng (2007), despite the writing task appearing to involve more involvement load. This limitation impedes the capacity to evaluate the involvement load hypothesis in relation to the acquisition of vocabulary through

speaking tasks. Nonetheless, further research in this area is warranted to explore the effectiveness of different types of tasks in improving vocabulary retention and production.

4.2 Research Questions

Thus, this study wanted to address the research question: Does adding a weekly 20-minutes speaking component (output) focusing on PV use to PV memorization with computer flashcards contribute to use more PVs in a speaking task, and to recall and recognize PVs measured by a gap-filling task?

4.3 Methodology

Participants and Groups

First, a priori power analysis using G*Power (Faul et al., 2007) with parameters using a one-tailed hypothesis for matched pairs with an estimated effect size of $d = 0.7$, $\alpha = .05$, and power of $1\text{-}\beta = .95$ indicated that the group size needed to be at least 24 for each of the three groups. Some attrition was anticipated due to the length of the study. Thus, I decided to recruit thirty participants per group.

During the ten-week memorization portion of this study, 111 Japanese university students majoring in English from two universities located in Fukuoka and Kyoto

prefectures participated.[7] The study consisted of one experimental group and two control groups. The experimental group (N = 39) and the second control group (N = 23) were recruited from University A, while the first control group was from University B (N = 49). The experimental participants were informed of the study, signed up in informed consent through *Google Forms*, and received compensation. The two control groups were composed of intact classes that did activities as part of their classes and later gave consent for their data to be used. The groups differed in being:

1. Memorization + Speaking group (Experimental condition) (N = 39): Studied the target PVs using computer flashcards at least three times a week. They spent 20 minutes a week producing language by telling picture-based stories that contained the target PVs, and listening to other participants do the same.

2. No Memorization + Speaking group (Control condition 1) (N = 49): Did not study the target PVs specifically but participated in the same picture-based story

[7] According to a *hensachi* (deviation score) of the universities based on the average score on the standardized Center Test (now called Unified University Entrance Examination), the Common European Framework of Reference for Languages (CEFR) level of university A was estimated to be between B1-B2, and that of university B was between B1-C1. It is worth noting that the test of English subject only assesses reading and listening skills.

speaking activities during a Speaking class. They also spent 20 minutes a week producing language by describing picture stories, and listening to other participants do the same. This group was used to determine the effects of explicitly studying the target PVs using computer flashcards for the Experimental group.

3. No Memorization + No Speaking (Control condition 2) (*N* = 23): Did not study the target PVs and did not participate in the picture-based story speaking activities. This control group was a Debate class, so they engaged in speaking activities during the semester, but did not focus on using any PVs. The class consisted of 90 minutes per session for 14 weeks, plus an introduction week.

Each participant in the groups where they participated in speaking activities was required to produce language for the specified task for 20 minutes per week by looking at pictures and telling a story that contained some of the target PVs intended to be used to describe the picture. All participants who engaged in the speaking activity received compensation as an incentive for their participation in the experiment.

Phrasal Verbs

This study used 45 PVs selected from the PHaVE List (Garnier & Schmitt, 2014), which provides the percentage of frequency for each meaning of polysemous PVs. Therefore, the most frequent meaning was selected to expose learners to the most important meanings in actual use. To prevent any bias from higher or lower frequency in a single *Anki* deck, the 150 PVs from the PHaVE List were divided into five sections of 30 PVs each. From each section, one PV was randomly selected for weekly use resulting in a total of five PVs studied each week, and all 45 PVs were studied over the nine-week experiment.

Procedure

All participants were given a weekly list of 5 PVs for a total of 9 weeks, with one additional blank week. They were asked to upload this list to *Anki* and study words at least 3 days a week. The flashcards in *Anki* included the English PV along with its Japanese translation and an example sentence in both English and Japanese (Figure 11). The target PVs in both sentences were highlighted in boldface. The practice of highlighting target vocabulary in sentences can draw learners' attention to them and motivate them to memorize them in context (Boers & Lindstromberg, 2012).

Figure 11. *Examples of Anki Decks*

Prior to the beginning of the experiment, the participants were provided with instructions on how to use *Anki*. Specifically, they were instructed to upload *Anki* decks using the "Basic (and reversed)" option for cards. This option presented the example sentences in English using targeted PVs and their corresponding Japanese translations, effectively doubling the number of cards for participants to study. The PVs were boldfaced in both the example sentences and translations to highlight their importance. To monitor their progress, the participants were required to submit a weekly PDF of their statistical data. This data included the number of days they had spent studying PVs. The participants who did not study their words for at least three days per week were excluded from the study. This exclusion criterion was implemented to ensure the accuracy and validity of the collected data.

For the storytelling, the speaking groups saw three pictures per week that incorporated all of the PVs from that week. All participants in the speaking groups who told stories were given one minute to read the description in Narrative Appendix B about the picture and look through the picture to assure that they understood the gist of the story, following the method of Derwing et al. (2006) study. The written description included the target PVs, but they were not underlined or highlighted in any way. This description was taken away before the participant began to describe the story. The participant who was describing the picture was encouraged to provide as many details as possible and to use at least two sentences for each frame, elaborating on aspects such as the weather, the character's location, activities, and emotions. This approach aimed to prevent the storyteller from concluding the narrative too quickly. The partner who was listening was asked to envision the story from the description and ask questions if any aspects were unclear.

Due to the spread of COVID-19, there were two speaking conditions: online and face-to-face. In the face-to-face speaking condition, each student was in charge of one of three stories and worked with two other students. They told the same story about two or three times to different listeners. In the online speaking condition, one participant described the first picture story (storyteller), and the other participant listened to the storytelling (listener). The listener could also indicate whenever they did not understand

something their partner was saying. After the storyteller finished describing the first picture, the listener took the turn to describe the second picture. Before the third story, the participants described the same story (first or second) one more time. After that, both participants saw the description of the third story and the picture and took their turns thinking up sentences. In both conditions, none of the participants knew which picture story they would be responsible for describing until the speaking activity began.

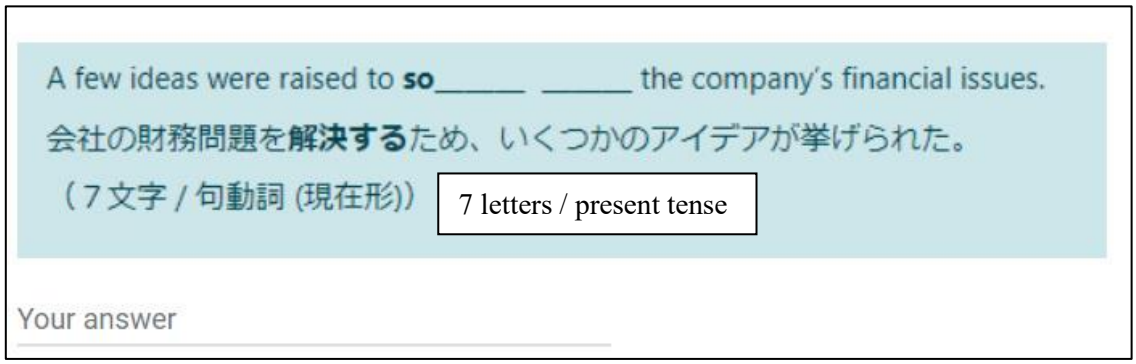
It is worth noting that many PVs have lexical equivalents or counterparts. For instance, 'postpone' can be considered a lexical equivalent for *put off* (for further explanation, please refer to *What Phrasal Verbs Are* in Section 2.1). During the study, participants were free to use lexical verbs in place of the target PVs. The primary focus of the study was to investigate the impact of the PV learning intervention on participants' usage of the target PVs and their proficiency in using them accurately and appropriately in tests.

Tests

This study conducted both written and speaking tests to examine the productive abilities of participants with PVs before and after a 10-week experimental period. The written test, implemented on *Google Forms*, assessed participants' proficiency in generating correct PVs in given sentences, encompassing a total of 45 targeted PVs.

The test featured sentences with blanks for the PVs, along with hints such as the first one or two letters, the tense of the target PVs, and the number of letters to be filled and a Japanese translation of the sentence. This test format was consistent with the one used in Experiment 1. To minimize the impact of context, which could lead to non-target PVs or expressions being considered as valid answers, the first two letters of the PV and a Japanese translation were offered as hints. However, for PVs with original verbs that were two or three letters long, then the first letter of the verb was provided as a hint along with the Japanese translation. The number of words and the verb tense in Japanese that needed to be filled in were displayed below the English sentence (Figure 12). To access a list of the test items, please consult Narrative Appendix C. For more comprehensive details, please refer to *Tests* in Section 2.3.

Figure 12. *Example of Gap-Filling Phrasal Verb Test*



The sentences of the pre-test and post-test were different. Sentences for the pre-test were drawn from the PHaVE List, while sentences for the post-test were sourced from Japanese-English and English-English online dictionaries such as, *Cambridge Dictionary*, *Collins English Dictionary*, *Oxford Learner's Dictionaries*, *Longman Dictionary of Contemporary English*, and *Alc*. Some of sentences were slightly changed to make them easier to understand. I provided Japanese translations for the sentences that had not been translated yet. The test was untimed but most students took about 20 minutes to complete the test. This test was modeled of the test used by Schmitt and Redwood (2011), and like their test, was piloted first with 21 native speakers and 11 non-native speakers to ensure validity and reliability. The native speakers made some mistakes so the test was adjusted and hints about the number of letters in the word and the tense of the word were added to make sure there was only one correct answer that native speakers would be able to guess from context. After revising and rewriting a few items this study administered the test to 3 more native speakers who all scored perfectly.

The study also included two speaking tasks to investigate whether memorization, output, and negotiation could help activate vocabulary for productive use. The first speaking task involved three multi-panel picture stories, each with two PVs that were intended to be used in the stories. The speaking task is similar to those that the speaking groups used every week. However, it differed from the tasks that the participants had

encountered previously. Each story of the test listed some words that were helpful for describing the story but did not list any of target PVs (see Narrative Appendix B for a description of the stories, Narrative Appendix D for the three stories used for the tests, and Figure 13 for the picture used in one of the three stories).

All participants were given two minutes to read the description of each story and then asked to describe the story in their own words. The second speaking task was an open speaking task that asked participants to describe what they would do in their ideal day for one minute. This task was not targeted toward the use of any specific PVs.

An immediate post-test was conducted two weeks after the last activity, with both the pre-test and post-test containing the same tasks (i.e., the same gap-filling task and speaking task in both tests).

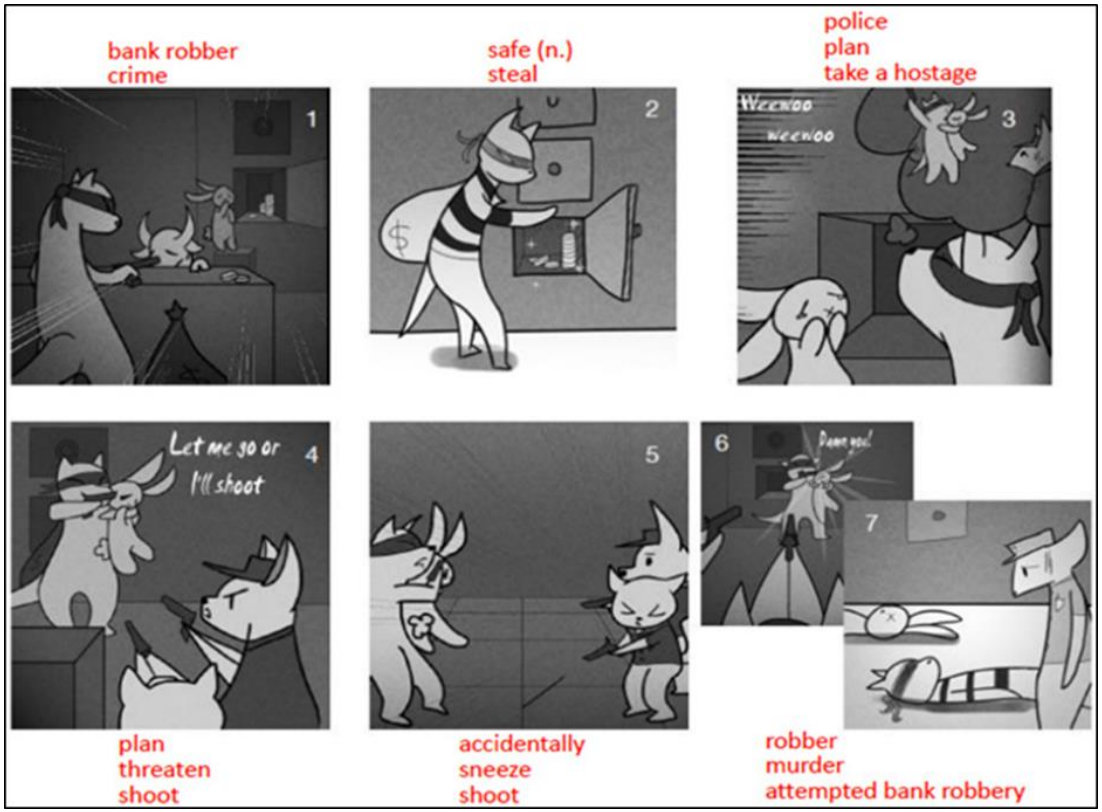
For analyzing the results of the gap-filling task, the gain score, which was the difference between the pre-test and post-test scores, was used. For the speaking task, the number of accurate PVs spontaneously produced by participants during the picture description task and spontaneous speaking task was counted. The gain score for speaking was then determined by subtracting the pre-test score from the post-test score.

To avoid ceiling effects, students who scored more than 80% on the initial written PV test were eliminated from the study.

First is the story as this study described it:

A very serious crime has happened at the bank. A robber took advantage of an open safe and tried to steal a lot of money. However, the police arrived and the robber faced a difficult situation. He came up with a plan to escape the police. He took a hostage! The robber tried to carry out his plan by threatening to shoot the hostage while he walked out of the bank. One of the policemen sneezed and accidentally shot his gun, so the robber committed murder and killed his hostage. The police immediately shot the robber so two people died in the attempted bank robbery.

Figure 13. *Example of Picture Description (Picture #1 in Pre-Test/Post-Test Material)*



It is worth noting that the control groups studied vocabulary lists on their vocabulary program at university that contained some PVs, including some non-targeted PVs with non-targeted meanings, such as *run away*, *show up*, and *work out* with non-targeted meaning.[8] Therefore, other target PVs that were not intended to be used in the picture description test but were studied through *Anki* were counted, such as *turn down*, *keep up with*, and *sit back*. However, PVs that were used incorrectly in students' speech, such as '*come up*' without '*with,*' were not included in the count.

Analysis

To access participants' proficiency in using PVs before and after the treatment, paired-samples *t*-tests were conducted for each of the two tasks separately.

In order to accurately assess the number of specific PV uttered, recordings were made in person or online using programs such as *Microsoft Teams* or *Zoom*. These recordings were then transcribed verbatim into a *Word* document. Transcriptions were made to accurately identify and count the number of correct PVs used by each

[8] *Work out* has more than two meaning. For example, 1) Plan, devise or think about STH carefully or in detail, 2) Exercise in order to improve health or strength (Garnier & Schmitt, 2014, in PHaVE List, p. 7). The present study used the first meaning in the list. However, some participants used *work out* as the second meaning in the second story. In this case, the PV was not counted because the meaning was not a targeted one.

participant for analysis purposes. Valid PVs were marked with a specific color, and the transcriptions were checked multiple times to ensure accuracy. These transcriptions are available upon request. The examples are as follows:

- Extracted from participant A in the first story: And he catches the rabbit and his plan taking a hostage is **carried out**.
- Extracted from participant B in the third story: And he began to uh- set up with- **set up** a table with many flowers and delicious cake.

The purpose of this analysis was to assess the participants' proficiency in using PVs accurately and to determine whether the intervention had a significant impact on their PV usage by calculating the number of correct PVs used by each participant.

4.4 Results

Prior to conducting any statistical analyses, the data were visually inspected in boxplots to assess their normality. Non-normal distributions were observed in all groups for the speaking test. In these cases, parametric bootstrap tests were used (Larson-Hall & Herrington, 2010). To determine whether there were any statistically significant differences between the three groups, a one-way analysis of variance (ANOVA) was performed. Effect sizes were also calculated to determine the magnitude of the

differences between groups. An alpha level of .05 was used to determine statistical significance.

Gap-Filling Phrasal Verb Test

The results of the gap-filling PV test where participants had to fill in a correct PV were analyzed. The average scores on the pre-test and post-test were calculated, and the gain scores from the pre-test to post-test were presented in Table 9.

Table 9. *Phrasal Verb Pre-Test and Post-Test Scores (Max. 45)*

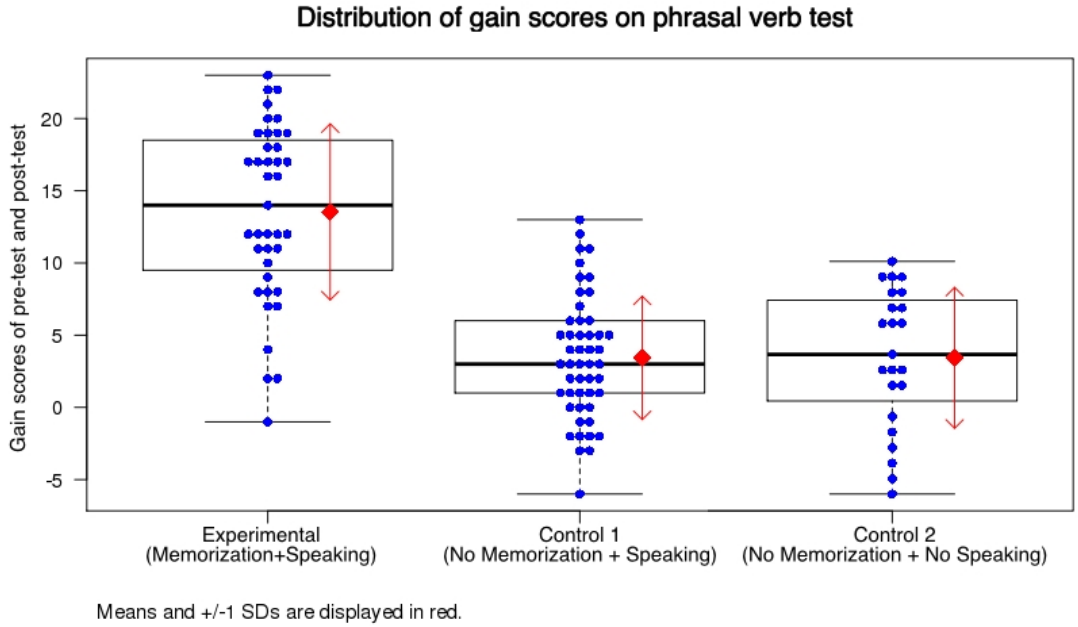
Groups	N	Pre-test Mean	Post-test Mean	Gain score	SD
Memorization + Speaking	39	20.79	34.33	13.54	6.10
No Memorization + Speaking	49	26.61	30.04	3.43	4.28
No Memorization + No Speaking	23	22.57	27.35	4.78	4.54

The Memorization + Speaking group, who studied target PVs using flashcards and retrieval practices, demonstrated the most significant improvement in their proficiency with an average gain score of 14 out of 45 possible points (M = 13.54, SD = 6.10). It is noteworthy that this group showed the most significant improvement in the study period, which lasted for 10 weeks. This finding suggests that the use of flashcards with retrieval practices was an effective method for enhancing PVs retention among

students. In contrast, the No Memorization + Speaking group, who described pictures without specifically studying PVs demonstrated a lower average gain score (M = 3.43, SD = 4.28). Similarly, the No Memorization + No Speaking group, who did not participate in any speaking activities, demonstrated a similarly small gain score to the No Memorization + Speaking group (M = 4.78, SD = 4.54). These results indicate that speaking activities alone may not be sufficient for improving PVs retention, and that learning PVs through flashcards with retrieval practices may be necessary for significant improvement.

The visualization in Figure 14 depicts a beeswarm plot, which displays the distribution of gain scores between the pre-test and post-test for the PV gap-filling test.

Figure 14. *Box Plot of Gain Scores on Phrasal Verb Gap-Filling Pre-Test and Post-Test*



The plot includes the median score, the 50% interquartile range, and the minimum and maximum scores, as well as marking any outliers in the data. The plot clearly illustrates that the experimental group had the highest average gain score compared to the control groups.

Furthermore, a one-way ANOVA test was conducted to determine if there were any significant differences in gain scores between the three groups. The analysis revealed a statistically significant difference between the groups ($F(2, 108) = 47.16$, $p < .0001$) (Table 10).

Table 10. *One-Way ANOVA Results for Phrasal Verb Test Gain Scores*

Source	Df	Sum Sq	Mean sq	F	Pr(>F)	η^2
Groups	2	2397.82	1198.91	47.16	0.0001	0.47
Residuals	108	2745.61	25.42			

Table 11 presents the results of the pairwise comparisons between the three groups' gain scores on the PV gap-filling test, along with their 95% confidence intervals (CIs) and effect sizes. The analysis found no statistical difference between the No Memorization + Speaking group and the No Memorization + No Speaking group, as their 95% CIs included zero.

However, there was a statistically significant difference between the Memorization + Speaking group and the No Memorization + Speaking group (95% CI: [7.54, 12.68]), and between the Memorization + Speaking group and the No Memorization + No Speaking group (95% CI: [5.61, 11.90]). The effect size of the Memorization + Speaking group and the No Memorization + Speaking group was larger (d = 1.92) than between the Memorization + Speaking group and the No Memorization + No Speaking group (d = 1.63), but both effect sizes were considered to be very large.

Table 11. *Pairwise Comparisons of Three Groups on Gap-Filling Phrasal Verb Test*

Groups	*t*	Pr(>|t|)	95% CI	
			Lower	Upper
Memorization + Speaking : No Memorization + Speaking	9.34	0.0001	7.54	12.68
Memorization + Speaking : No Memorization + No Speaking	6.61	0.0001	5.61	11.90
Memorization + No Speaking : No Memorization + No Speaking	1.06	0.54	-4.38	1.67

These results indicate that studying PVs using the flashcards and doing speaking activities using the PVs were more effective in improving PV production in the gap-filling post-test compared to the other two groups who did not specifically study PVs.

Speaking Test

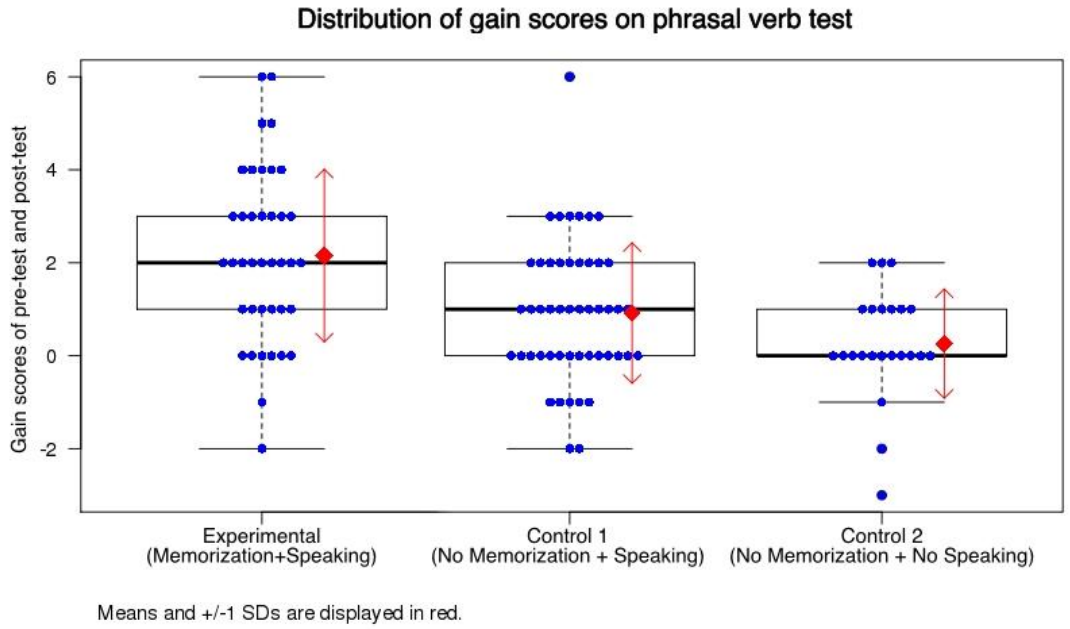
The results of the speaking test, where participants were asked to describe picture stories and their ideal day, were analyzed by calculating the number of PVs used by each participant. The average number of correct PVs used by each participant was calculated for both the pre-test and post-test. The cumulative number of correct PVs used in all four speaking tasks is presented in Table 12 and visually represented in Figure 15.

The experimental group demonstrated the largest increase in the number of target PVs uttered from pre-test to post-test (M = 2.15, SD = 1.86). It should be noted that there is no established upper limit for the number of the target PVs that can be used in describing the given picture stories and ideal day scenario.

Table 12. *Phrasal Verb Production on Speaking Tests*

Groups	N	Pre-test Mean	Post-test Mean	Gain score	SD
Memorization + Speaking	39	0.90	3.13	2.15	1.86
No Memorization + Speaking	49	1.43	2.32	0.91	1.51
No Memorization + No Speaking	23	0.61	0.91	0.26	1.18

Figure 15. *Box Plot of Gain Scores on Speaking Test*



On the other hand, the No Memorization + Speaking group, who did not study PVs but participated in speaking activities, showed medium-size progress in PV production from the pre-test to post-test (M = 0.91, SD = 1.51). Some students in the group may have learned some specific PVs during speaking, even though they were not encouraged to use them.

In contrast, the No Memorization + No Speaking group, who did not participate in any speaking activities, demonstrated the smallest increase in the number of target PVs used (M = 0.26, SD = 1.18).

Moreover, a one-way ANOVA was conducted to examine whether there were significant differences in the gain scores off the three groups on the number of PVs

produced in the speaking tests. Analysis of the data, as presented in Table 13, revealed a statistically significant difference among the groups ($F(2, 108) = 11.87$, $p < .0001$). This indicates that there were variations in the number of PVs produced by the groups during the speaking tests.

Table 13. *One-Way ANOVA Results for Gain Scores on Speaking Test*

Source	Df	Sum Sq	Mean sq	F	Pr(>F)	η^2
Groups	2	59.63	29.813	11.87	0.0001	0.18
Residuals	108	271.19	2.511			

Following the ANOVA analysis, pairwise comparisons were conducted to determine the statistical differences between the three groups on PV production in the speaking tests.

The results, as presented in Table 14, indicate that there were no statistically significant differences between the No Memorization + Speaking group and the No Memorization + No Speaking group, with a 95% confidence interval (95% CI: [-0.29, 1.61]). This implies that engaging in speaking activities alone without specific focus on PVs did not lead to any significant improvement in PV production.

Table 14. *Pairwise Comparisons of Three Groups on Speaking Test*

Groups	*t*	Pr(>|t|)	95% CI	
			Lower	Upper
Memorization + Speaking : No Memorization + Speaking	3.63	0.001	1.24	0.43
Memorization + Speaking : No Memorization + No Speaking	4.54	0.001	0.90	2.88
Memorization + No Speaking : No Memorization + No Speaking	1.64	0.23	-0.29	1.61

However, there was a statistical difference between the Memorization + Speaking group and the No Memorization + Speaking group, with a 95% CI [1.24, 0.43] indicating that the difference between the two groups was small. This suggests that studying PVs using flashcards and engaging in speaking activities using PVs had a positive impact on PV production.

Moreover, the comparison between the Memorization + Speaking group and the No Memorization + No Speaking group revealed a statistically significant difference, with the 95% CI ranging from 0.90 to 2.88. This suggests that the Memorization + Speaking group performed significantly better than the No Memorization + No Speaking group in terms of PV production in speaking.

The effect size of the Memorization + Speaking group and the No Memorization + No Speaking group was found to be $d = 1.21$, indicating a large effect size. On the

other hand, the effect size of the Memorization + Speaking group and the No Memorization + Speaking group was found to be $d = 0.73$, indicating a medium effect size. Therefore, while there was still a statistically significant difference between the Memorization + Speaking group and the No Memorization + Speaking group, the effect size was not as large as that between the Memorization + Speaking group and the No Memorization + No Speaking group.

However, upon observing Figure 15, there were some outliers in the three conditions, indicated by dots beyond the box plot whiskers. In order to check if the data followed a normal distribution, a Shapiro-Wilk normality test was conducted. The results of the test indicated that the data did not adhere to a normal distribution, which is an important consideration when performing hypothesis testing. Specifically, the experimental group exhibited non-normality ($W = 0.97$, $p = .38$), with a p-value exceeding the significance level of 0.05. Conversely, the two control groups demonstrated normality ($W = 0.93$, $p < .01$ for the No Memorization + Speaking group; $W = 0.86$, $p < .01$ for the No Memorization + No Speaking group). To overcome this issue, a bias-corrected and accelerated bootstrap was conducted using Wilcox's (2012) Weighted Random Sampling (WRS) method (as cited in online chapter "Bootstrapped and Means-Trimmed One-Way ANOVA and Multiple Comparisons In R," Larson-Hall, 2015, p. 331) in R. WRS is used when the assumptions of normality and homogeneity

of variance are not met. The results of hypothesis testing were given for pair comparisons with 20% trimmed means.

Table 15 indicates that all group comparisons are statistically significant. There was a statistically significant difference between the Memorization + Speaking group and the No Memorization + Speaking group (p = .01, 95% CI: [-2.23, -0.34]), the Memorization + Speaking group and the No Memorization + No Speaking group (p = .01, 95% CI: [-2.69, -0.88]).

Table 15. *Bootstrapped Pairwise Comparisons of Three Groups on Speaking Tests*

Groups	Test value	Standard error	p value	95% CI	
				Lower	Upper
Memorization + Speaking : No Memorization + Speaking	-3.07	0.42	0.01	-2.23	-0.34
Memorization + Speaking : No Memorization + No Speaking	-4.46	0.40	0.01	-2.69	-0.88
Memorization + No Speaking : No Memorization + No Speaking	1.98	0.26	0.04	-0.07	1.08

The p-values for these comparisons are below the significance level of .05, and the 95% CIs do not include zero. Therefore, the null hypothesis was rejected for all comparisons, suggesting that the observed results are statistically significant and not likely to have occurred by chance. However, in the comparison between the Memorization + No

Speaking group and the No Memorization + No Speaking group (p = 0.04, 95% CI: [-0.07, 1.08]), statistical significance was not found. This is because the 95% CI includes zero, even though the p-value is less than the significance level of .05. Thus, the null hypothesis for this specific comparison was not rejected.

4.5 Conclusion

The present study aimed to investigate the effects of memorization and speaking on the acquisition and retention of PVs among English learners at a university in Japan. The results indicated that both memorization and speaking contributed significantly to the improvement of PVs knowledge among the participants. Moreover, the study showed that the combination of memorization and speaking was the most effective approach to enhance the learners' PVs performance. Specifically, the Memorization + Speaking group outperformed the other two groups in all measures of PVs acquisition and retention, as demonstrated by the significant gain scores observed in the gap-filling test.

Educators can use the results to design more effective PVs teaching materials and activities that incorporate both memorization and speaking. Furthermore, the study highlights the importance of engaging learners in speaking practice to enhance their PVs knowledge and retention. This finding is particularly relevant for EFL learners, as they may have limited exposure to spoken English in their daily lives.

4.6 Discussion

The present study emphasizes the essential role of a combined approach involving memorization and output activities for effective PV learning. Memorization through spaced retrieval practice plays a significant role in PV learning by making the PVs readily available for use during speaking activities. Furthermore, the study highlights the importance of output activities, with a focus on speaking tasks, in PV learning. Engaging in speaking activities provides learners with opportunities to actively use and practice the PVs they have memorized. Additionally, engaging actively is crucial in effective vocabulary learning as it involves exploring how words are used in different contexts and understanding their relationships with other words (Stahl & Kapinus, 2001). This aligns with the notion, highlighted in the present study, that engaging in speaking activities allows learners to actively practice and apply PVs, thereby deepening their understanding and retention.

This study's findings are consistent with previous research that emphasizes the importance of memorization using spaced retrieval practice (Chapter 2) and collaborative output activities (e.g., Kowal & Swain, 1994; Kuiken & Vedder, 2002; Leeser, 2004; Nassaji & Tian, 2010; Teng, 2017) in language learning. While there is limited research specifically investigating the combination of memorization and output activities, collaborative output activities have shown beneficial effects in PV learning.

These activities facilitate learners in identifying gaps in their language knowledge, directing their attention towards the connection between form and meaning, and receiving valuable feedback from their peers.

I recommend that language educators incorporate both memorization and interactive output activities, to enhance PV learning. By incorporating collaborative speaking tasks and deliberate memorization exercises, students can experience a comprehensive learning approach that promotes PV acquisition. This recommendation is supported by closely related studies on vocabulary acquisition, including the research conducted by Hulstijn and Laufer (2001) and Laufer and Hulstijn (2001) on the involvement load hypothesis. These studies emphasize the significance of active involvement with words and meaningful practice in vocabulary learning. Taking these findings into account further reinforces the pedagogical suggestion that a combination of memorization and output activities is crucial for successful PV learning.

Further research is needed to address some limitations of this study. One potential area of improvement is to include an additional comparison group that studies the target PVs using flashcards but does not engage in speaking activities. This would help determine if students can effectively utilize their word knowledge without output activities.

Moreover, it is important to acknowledge that the present study had a limited duration of one academic semester, which may restrict our understanding of the long-term retention of PVs. While the participants in the experimental group successfully memorized and utilized the vocabulary during the study period, it remains uncertain how well they would retain this knowledge six months or a year after completing the study. Unfortunately, due to time and budget constraints, the study was unable to investigate this aspect. Additionally, the study did not include variations in the illustrations or test formats. It is worth noting that when the same test is administered multiple times and test takers remember the content, it raises concerns about practice effects or test familiarity. This can impact the validity and reliability of the test results, as participants may recall specific answers from previous attempts rather than demonstrating genuine knowledge or learning. Consequently, future studies should address this issue by implementing different versions of the test or employing counterbalancing techniques to minimize the influence of test familiarity.

Addressing these limitations through further research will enhance our understanding of PV learning and contribute to the development of effective instructional strategies.

CHAPTER 5.

Experiment 4: Speaking Fluency with the Use of Collocations

Language acquisition is a fascinating process that has captivated the attention of researchers for decades. As one of those language learners, I find it interesting to observe the different methods that learners use to acquire new words and phrases and how they can use them in a native-like way. One area of language learning that has recently gained more attention is the acquisition of formulaic sequences (FSs), which are pre-built, commonly used word combinations that are essential to speaking fluency.

Despite the importance of FSs, learners often struggle to produce them in actual situations, and it can be challenging to acquire native-like speaking fluency. This is where vocabulary programs come in as they offer a quick and effective way for learners to pick up new words. However, while list learning is useful, it is not enough on its own. Learners need to be able to use the vocabulary they are learning in real-life situations if they want to communicate effectively.

In the previous chapter, I explored the challenges that language learners face in mastering FSs, with a specific focus on phrasal verbs (PVs), in their spoken language. I understand how difficult it can be for learners to understand the multiple and figurative meanings of PVs and use them effectively in speaking situations, observing the limited use of PVs in their utterances. This exploration prompted me to examine whether the

inclusion of focused on a different type of FSs, specifically collocations, within a speaking component could improve learners' speaking fluency. Collocations are word combinations that frequently occur together, such as 'take note' or 'significant decrease,' and are more straightforward in structure with more predictable meanings.

In Chapter 5, the main objective of this study was to examine the potential impact of integrating a speaking component into collocation memorization through computer flashcards, with a specific focus on enhancing learners' speaking fluency. By exploring the potential benefits of incorporating speaking practice into collocation instruction, this study aimed to provide valuable insights into how learners can move beyond mere acquisition to actively using FSs in their communication.

At my university, mid- to upper-intermediate level students were using a vocabulary program to learn academic vocabulary. However, I sought to take their learning a step further by incorporating a speaking component, specifically focusing on the learning of verb-noun and adjective-noun collocates, which are a type of FSs. My hope was to assist L1 Japanese users of English in improving their speaking fluency and effectively integrating new vocabulary into their daily communication.

5.1 Literature Review

Formulaic Sequences: Collocations

FSs have different types of expressions, such as idioms, phrasal verbs, lexical bundles, and collocations. Among these expressions, research on the acquisition of collocations is still in its early stages, interest in this area has been growing (Biber, 2009). Identifying collocations is a crucial initial step in studying them, and this task has been facilitated by the use of corpus approaches, which offer a vast database of real language usage for analysis.

Nguyen and Webb (2017) created a collocation test using headwords from the 3,000 most frequent words of English in order to measure Vietnamese EFL learners' receptive knowledge of collocation. They defined collocations as "any two words in combination that emerge from a corpus at a greater frequency than could occur by chance" (2017, p. 3), but limited the collocations they studied to verb-noun and adjective-noun combinations. Nguyen and Webb's (2017) list of 180 collocations, 30 each of verb-noun and adjective-noun collocations at each level of frequency (1,000, 2,000, and 3,000 words). Their list was created by randomly choosing nouns, identifying collocations with a span of minus one (one word to the left of the node words), finding only collocates where the collocating word was the same or lower frequency than the node noun, and choosing only collocations with a frequency of at

least 50 in *Corpus of Contemporary American English* (COCA) (Davies, 2008-) and a mutual information score between the node and collocate of at least 3.0.

In testing these collocations, Nguyen and Webb (2017) used a four-item multiple choice test. They found that knowledge of the correct collocation was quite low; scores were less than 50% correct for both verb-noun and adjective-noun collocates. The participants did know more of the higher-frequency collocates than the less-frequent ones.

Since LX learners' knowledge of collocations is limited, Nguyen and Webb's (2017) findings underscore the need for collocations to be implemented into vocabulary learning programs (for the term 'LX,' see Footnote 1). This study thus employed the same collocations that would be useful for learners to learn and improve their speaking fluency.

Speaking Fluency with Formulaic Sequences

There are two main approaches that have been proposed to address how vocabulary should be learned, namely deliberate learning or incidental learning. However, deliberate learning is more time-efficient than incidental learning (Horst, 2005; Hulstijn, 2001). It seems that learning vocabulary deliberately is more efficient and effective for time-limited university classes.

There are two studies that I am aware of that address deliberate learning of FSs. The first study conducted by Taguchi (2007) looked at how speaking fluency is influenced through the acquisition of FSs in speaking. In her study, the participants were Japanese beginning learners of English that is different from the present study, but her study looked into their speaking fluency using multi-word sequences which is relevant to the present study. According to Taguchi (2007), practicing and memorizing grammatical chunks resulted in improvement in Japanese learners' fluency in spontaneous conversation with situational descriptions and narrative tasks describing a third person. There was a strong correspondence between the frequency of the chunks that the learners encountered in their study and their production of these chunks spontaneously in conversation. Taguchi (2007) provides strong evidence that targeted practice of specific multi-word sequences can enhance their integration into spoken language. However, the chunks studied in Taguchi (2007) were very basic, beginning-level chunks such as "let's X" and question words. These are more like grammatical chunks than the less frequent collocations this study was targeting such as '*work hard*' and '*natural beauty*.' I thus believed it is an open question whether studying specific words leads to them being used in speaking, and use of these specific words leads to better fluency.

The other study which looked at how speaking fluency is influenced through the acquisition of FSs in speaking was McGuire and Larson-Hall (2017). They conducted a five-week study of ESL students studying in the U.S. The study had only a small number of participants (control group $N = 8$, experimental group $N = 11$). Participants produced dialogues pre- and post-instruction. The instruction in both groups used a task-based approach and both groups saw the same transcripts of spoken tasks. The teacher of the experimental group aimed to raise awareness of FSs but did not explicitly direct participants to memorize any particular words, while the control group was aware of isolated vocabulary and grammar. McGuire and Larson-Hall (2017) calculated the ratio of syllables containing FSs to all syllables in the pre-test and post-test. They then reported that participants in the experimental group increased FSs in their speaking from 25% to 35%, whilst the control group dropped from 29% to 27%. A paired-samples *t*-test on speech rate and mean length of run found that there was no statistical difference from pre-test to post-test for participants in the control condition (effect sizes of $d = 0.06$ were obtained for speech rate, and $d = 0.17$ for mean length of run) while the experimental condition showed statistical differences with notably large effect sizes ($d = 1.3$ and $d = 1.1$). This means that participants in the control condition did not speak statistically any faster after the treatment than they had before, while the experimental condition spoke much faster after the treatment. This study suggests that explicit

instruction of FSs can improve speaking fluency even when specific FSs are not targeted. However, due to its small sample size and non-standard interpretation of FSs, broad claims cannot be made. My research study will apply this type of analysis to students who have explicitly studied particular FSs and have larger group sizes.

The above studies focused on deliberate learning of FSs. Other studies on incidental FS learning have failed to demonstrate significant statistical improvements in LX learners' speaking fluency (e.g., Boers et al., 2006; Stengers et al., 2010), although some progress was observed. Thus, it was thought that a control group, who would not study collocations deliberately in the present study, might show some small gains in their speaking fluency. However, I expected that the gains would be significantly different from the experimental group.

What Kind of Speaking Task might be Useful for Speaking Fluency

The importance of helping learners notice gaps between their intended speech and their actual ability was discussed in Chapter 4. This aspect will not be further explored in this literature review. For a comprehensive analysis of related studies, please refer to the literature review titled *Collaborative Learning on Speaking* in Section 4.1, where the

subject matter is thoroughly examined. Consequently, the present study has chosen to incorporate paired learners engaging in collaborative speaking activities.

5.2 Research Questions

Thus, this study wanted to address the research question: Does the implementation of flashcard learning combined with a speaking component focused on collocations improve learners' speaking fluency?

5.3 Methodology

It should be noted that the methodology discussed in this chapter shares similarities with the methodology presented in Chapter 4. However, to avoid repetition, the overlapping parts will not be repeated here. For more detailed information, please refer to *Methodology* in Section 4.3.

Participants and Groups

Firstly, a priori power analysis was conducted using G*Power (Faul et al., 2007) with a one-tailed hypothesis for matched pairs and an estimated effect size of $d = 0.7$, $\alpha = .05$, and power of $1\text{-}\beta = .95$. The analysis indicated that at least 24 participants were needed

for each of the three groups, considering some attrition was expected due to the long-term nature of the study. Therefore, more than 24 participants were recruited per group.

A total of 73 students majoring in English from two Japanese universities located in Fukuoka and Kyoto prefectures were recruited for the study. The experimental group (N = 51) was independently recruited from University A and B, which were the same universities as in Chapter 4. Participants in the experimental group provided informed consent by filling out a *Google Forms* and received compensation for their participation. On the other hand, the control group (N = 22) consisted of intact classes from University A, who performed activities as part of their classes. The participants in the control group gave consent for their data to be used. It is important to note that the experimental and control groups received different treatments, as follows:

1. Experimental group: Studied collocations at least three times a week and spent 20 minutes a week describing and listening to other participants describe picture stories.

2. Control group: Did not study the target collocations and did not participate in the picture story speaking activities. This control group was a Debate class, so they engaged in discussion activities during the semester, but did not focus on using

any collocations nor participate in picture description activities. The class consisted of 90 minutes per session for 14 weeks, plus an introduction week.

The story contained some collocations intended to be used to describe the picture. All participants who participated in the speaking activity received compensation as an incentive to complete the experiment.

Procedure

The participants in the experimental group learned 100 collocations for ten weeks. The complete list of collocations used in this study can be found in Narrative Appendix E. To facilitate their learning and preparation for speaking activities, the participants were instructed to study 10 collocations each week using the *Anki* flashcard program. They were encouraged to review the collocations at least three times a week. This assumption is built upon the idea that students would be able to memorize FSs, particularly PVs, as indicated by the findings in Chapters 2 and 4. The primary objective, however, was to assess whether the participants would employ these collocations in spontaneous speaking during the post-test and whether this usage would contribute to enhanced fluency in their speech.

During the experimental phase, the participants in the experimental group were exposed to three pictures per week. Their task was to comprehend a picture story and then retell it using their own words. Before describing the picture, they were provided with one minute to read the description given in Narrative Appendix F and examine the picture to ensure they understood the main idea or gist of the story. This approach aimed to promote comprehension and enhance their ability to effectively retell the story based on their understanding of the picture.

In the speaking activities, each participant in the experimental group was assigned one of the three stories, and they worked together with two other participants. They took turns telling the same story to two different listeners, repeating the storytelling process two or three times. The assignment of stories to participants was randomly done each week, just before the speaking activity took place.

Tests

To assess the participants' speaking fluency, tests were administered both before and after the 10-week experiment. The test consisted of three stories, and all participants were tested on the same three stories. For each story, a list of words that could assist in describing the story was provided. However, the list did not include any of the targeted collocations (each story had four specific collocations). For the detailed stories used in

the test, please refer to Narrative Appendix G, and for the associated picture for one of the stories, please consult Figure 13 in Section 4.3.

Collocations

The study selected 100 collocations from a list of 180 provided by Nguyen and Webb (2017). The collocations were chosen based on their suitability for illustration in picture stories, with a focus on those that were easier to understand. The list included 30 verb-noun and 30 adjective-noun collocations at each of the three frequency levels (1,000, 2,000, and 3,000 words). Out of these 100 collocations, the study used 23 verb-noun collocations from the 1,000-word level, 21 from the 2,000-word level, and 6 from the 3,000-word level. In addition, 23 adjective-noun collocations were chosen from the 1,000-word level, 17 from the 2,000-word level, and 10 from the 3,000-word level.

Fluency Analysis

To analyze speaking fluency, the study used objective measurements of temporal variables, specifically speech rate (SR) and mean length of run (MLR). SR is calculated by dividing the number of syllables by the length of time, including both the learner's utterances and pauses. For example, if a participant speaks for 30 seconds and produces 60 syllables, their speech rate would be 2 syllables per second (60 syllables / 30

seconds). MLR is the average number of syllables spoken between pauses. For example, if a participant speaks for 1 minute, with a total of 10 pauses and 30 syllables between pauses, their MLR would be 3 syllables per pause (30 syllables / 10 pauses).[9]

Kormos and Dénes (2004) found that the best measurements for fluency are speech rate, mean length of utterance, phonation time ratio, and the number of stressed words produced per minute. However, the present study chose to use SR and MLR, based on a similar study by Wood (2009) that used these variables to measure speaking fluency in a Japanese learner of English who was taught formulaic sequences and asked to use them in speaking practice.

5.4 Results

The study conducted statistical tests to determine whether there were any significant differences between the experimental group and the control group in terms of their

[9] In terms of SR and MLR, which take syllable count into account, it is important to consider the potential difference between collocations and lexical items. For instance, collocations like 'senior citizens' are likely to have more syllables compared to lexical items like 'the elderly,' which could impact the analysis. In this study, the analysis of SR and MLR aimed to investigate the impact of incorporating a speaking component with collocations on learners' speaking fluency. By examining these temporal variables, the study aimed to gain insights into how the intervention influenced the pace and flow of learners' speech.

speech rate (SR) and mean length of run (MLR) gains on two tasks: descriptive storytelling and spontaneous speech. To determine whether there were any statistically significant differences between the two groups, an independent samples *t*-test was performed. Effect sizes were also reported to examine the magnitude of the differences between the groups, and an alpha level of .05 was used to determine statistical significance. the data were examined for normality using the Shapiro-Wilk test to determine whether the assumption of normality was met. Non-normal distributions were found in the control group for SR on the descriptive storytelling task and in the experimental group for MLR on the same task. Therefore, parametric bootstrap tests were used in those cases (Larson-Hall & Herrington, 2010).

Speech Rate

The speech rate was measured in syllables per second on the descriptive storytelling task where participants described three stories and spontaneous speech task. Table 16 shows the descriptive statistics of the SR gains on both tasks for the experimental and control groups.

The experimental group gained their SR on the descriptive storytelling task (M = 0.30, SD = 0.29), while the control group lost their SR from the pre-test to post-test (M = -0.75, SD = 0.53). However, on the spontaneous speech task where participants talked

about their ideal day for one minute, the control group gained their SR (M = 0.56, SD = 0.39) more than the experimental group (M = 0.21, SD = 0.35). This is surprising because the control group was not instructed to study specific collocations but showed some improvements in SR on the spontaneous speech.

Table 16. *Speech Rate (Syllables per Second)*

Groups	N	Speaking tasks	Pre-test Mean	Post-test Mean	Gain	SD
Experimental	51	Descriptive storytelling	1.86	2.16	0.30	0.29
		Spontaneous speech	2.17	2.38	0.21	0.35
Control	22	Descriptive storytelling	1.72	0.97	-0.75	0.53
		Spontaneous speech	2.08	2.14	0.56	0.39

Figure 16 and Figure 17 show the distributions of each participant's gain score between the pre-test and post-test for the SR of both tasks. It is evident that most participants in the control group lost gains in their SR on the descriptive storytelling task.

An independent samples t-test showed that no statistically significant differences between groups were found in how quickly the participants spoke on the spontaneous speech task ($t(71) = -1.70$, $p = .09$). However, for the descriptive storytelling task, there

were statistically significant differences between the groups ($t(71) = -10.97$, $p < .001$) since the p-value was much less than the .05 level. The 95% confidence interval (CI) showed no statistical difference between the experimental and control group (95% CI: [-0.34, 0.27]) on the spontaneous speech task. However, there was a statistical difference between the experimental and control (95% CI: [1.24, 0.86]) on the descriptive storytelling task. The effect size of the control and experimental group on SR of the descriptive storytelling ($d = 2.46$) was much larger than the spontaneous speech task ($d = 0.94$).

Moreover, a Shapiro-Wilk normality test found non-normality on the control group ($W = 0.91$, $p = .04$) since the p-value was less than .05 level. A 20% trimmed mean was calculated using Yuen code (as cited in a vignette of Mair & Wilcox, 2020, p. 7) in R. An independent samples t-test showed there was a statistical difference between groups ($t(15.78) = 8.79$, $p < .01$, 95% CI: [-1.43, -0.87], Cohen's $d = 0.89$).

Figure 16. *Speech Rate for Descriptive Storytelling*

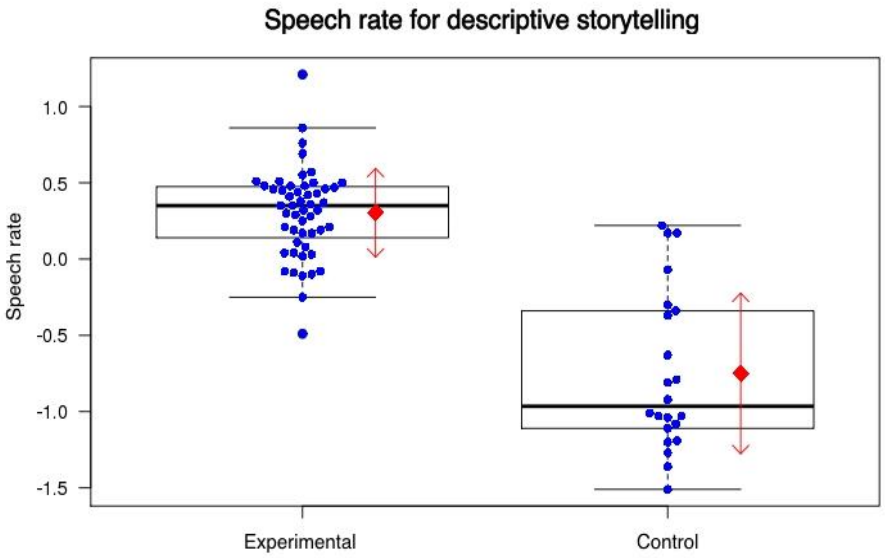
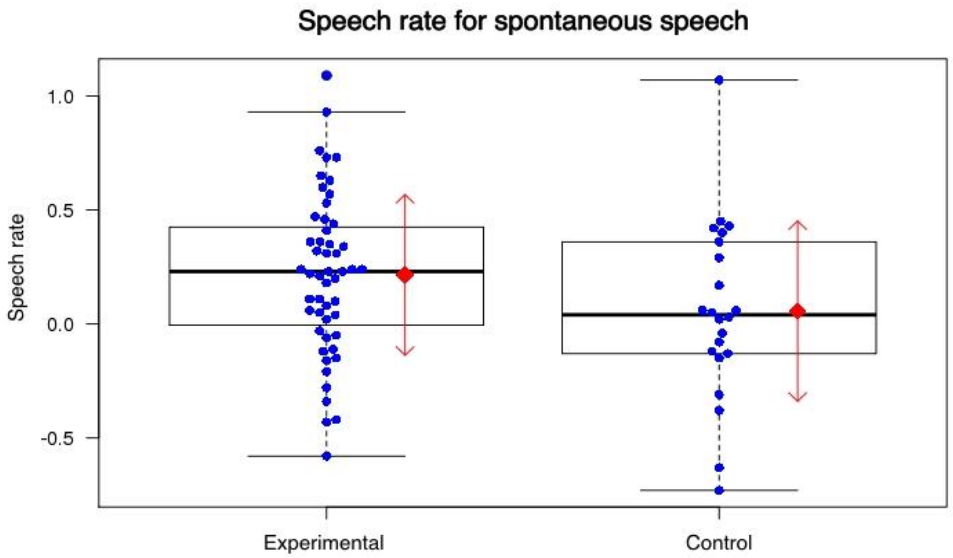


Figure 17. *Speech Rate for Spontaneous Speech*



Mean Length of Run

In terms of mean length of run (MLR), which is the average number of syllables participants uttered between pauses, Table 17 shows descriptive statistics about MLR gains on both tasks. On the descriptive storytelling task, the average gain of MLR for the experimental group gained was 0.27 while the control group experienced a significant decrease in MLR gain from the pre-test to post-test, -3.35. On the spontaneous speech task, the experimental group's MLR gain was 0.67 while the control group's MLR gain was 0.21.

Table 17. *Mean Length of Run*

Groups	N	Speaking tasks	Pre-test Mean	Post-test Mean	Gain	SD
Experimental	51	Descriptive storytelling	4.16	4.43	0.27	1.04
		Spontaneous speech	4.58	5.25	0.67	1.35
Control	22	Descriptive storytelling	3.81	0.45	-3.35	0.61
		Spontaneous speech	4.55	4.76	0.21	1.07

Figure 18 and Figure 19 illustrate the distributions of each participant's gain scores between the pre-test and post-test for the MLR of both tasks. The MLR gains for

the spontaneous speech task spread over the graph regardless of the groups as shown in Figure 19. The results of both tasks are similar to the distributions of SR.

The results of an independent samples *t*-test result showed there was a significant difference between the pre-test and post-test on MLR of descriptive storytelling ($t(71) = -15.22$, $p < .001$, 95% CI: [4.10, 3.15]). However, there was no statistical difference between the tests for MLR of spontaneous speech ($t(71) = -1.41$, $p = .16$, 95% CI: [-1.11, 0.19]). The effect size of MLR on the descriptive storytelling was large (Cohen's $d = 4.25$) for the control and experimental group. On the other hand, the effect size on the spontaneous speech task was small-to-medium (Cohen's $d = 0.38$) for the two groups.

It is worth noting that several outliers were observed in the experimental group for the descriptive storytelling task, as shown in Figure 18 (outliers beyond the minimum/maximum whiskers of the boxplot). Additionally, a Shapiro-Wilk normality test indicated non-normality in the control group ($W = 0.91$, $p < .01$) since the *p*-value was less than the alpha level of .05. A 20% trimmed mean was calculated using Yuen code (as cited in a vignette by Mair & Wilcox, 2020, p. 7) in R. Finally, an independent samples *t*-test showed there was a statistical difference between groups ($t(27.63) = 24.68$, $p < .001$, 95% CI: [-3.90, -3.30], Cohen's $d = 0.95$).

Figure 18. *Mean Length of Run for Descriptive Storytelling*

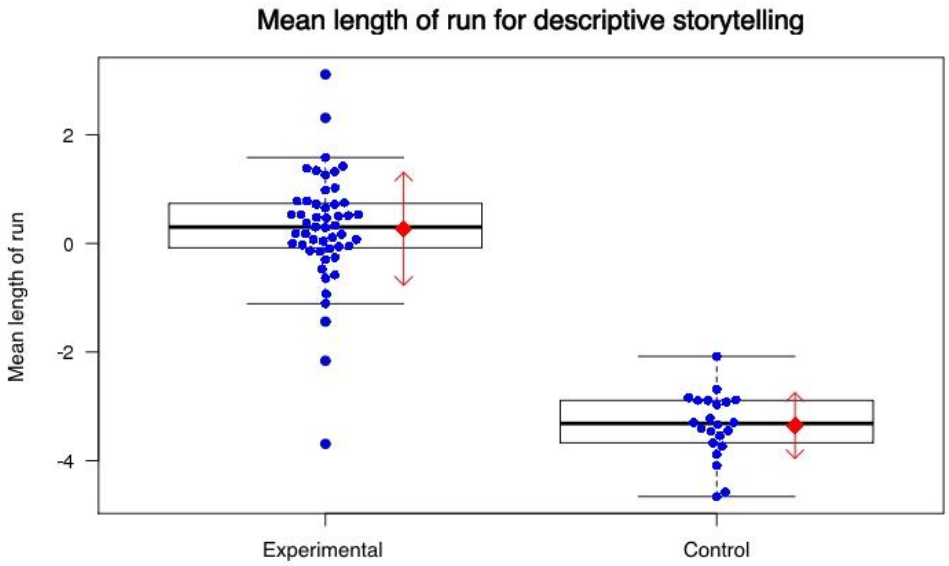
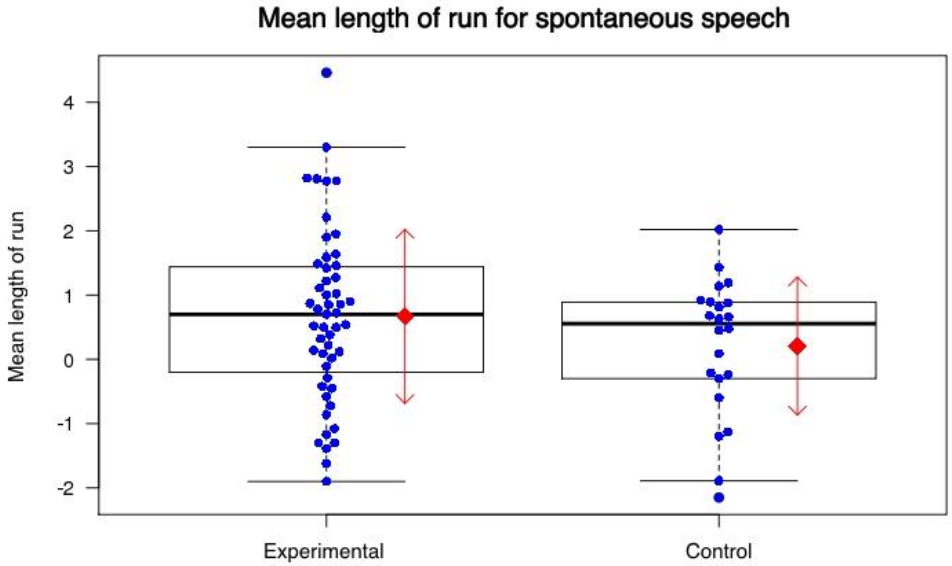


Figure 19. *Mean Length of Run for Spontaneous Speech*



Overall, the study found that the experimental group showed significant improvements in speech rate on the descriptive storytelling task and MLR on both tasks compared to the control group. The study could not affirm why the SR of the control group improved on the spontaneous speech task in a way similar to the experimental group. The findings suggest that collocation instruction can be beneficial for students to improve their speech fluency and can lead to better performance on tasks that require more narrative skills. However, further research is necessary to explore the extent to which collocation instruction can enhance other aspects of oral proficiency.

5.5 Conclusion

The study aimed to investigate the effectiveness of a collocation instruction intervention in enhancing the SR and MLR of Japanese university learners of English, specifically focusing on descriptive storytelling and spontaneous speech tasks.

The results pertaining to MLR on the descriptive storytelling task indicated no significant effect, except for a noteworthy increase in pauses (corresponding to a decrease in MLR) observed in the control group at the post-test. A similar pattern observed for SR, with the control group exhibiting a considerable decrease in SR during the post-test for the descriptive storytelling task. However, in the case of the experimental group, there was a modest to moderate gain in SR specifically for the

descriptive storytelling task. These findings imply that the acquisition of collocations has the potential to enhance learners' speaking fluency, as indicated by changes in SR.

Conversely, in the case of the spontaneous speech task, no statistically significant differences were observed in both SR and MLR between the experimental and control groups. Although the experimental group demonstrated improvements in SR and MLR scores, comparable to those of the control group, the reasons underlying the control group's improvement remain unclear based on the present study's findings.

In summary, the study demonstrated that collocation instruction using flashcards and picture description activities to teach collocations can be effective in improving the SR and MLR of Japanese learners on specific speaking tasks.

5.6 Discussion

The observed gains in both SR and MLR among the control group for the spontaneous speech task raise questions that warrant further investigation. Specifically, it is crucial to explore the underlying reasons for the improvement in speaking fluency among the control group who neither engaged in collocation instruction nor participated in any specific speaking activities.

One potential direction for investigation is to explore whether the observed improvement in speaking fluency among the control group resulted from incidental

acquisition through exposure to English language input in their environment, such as university classes or other unidentified factors. Research suggests that incidental vocabulary acquisition requires a substantial number of encounters. For instance, studies on graded reading indicate that it may take over 20 encounters for learners to incidentally acquire new vocabulary (Waring & Takaki, 2003). Additionally, it has been found that the repetition effect is less pronounced in listening compared to reading, even with 20 encounters of the target words (Brown et al., 2008). However, considering the participants in the present study likely did not encounter the target collocations over 20 times in their EFL environment, other factors may have contributed to their enhanced fluency.

On one hand, Saito and Hanzawa (2016) suggest that increased attendance in English classes and higher usage of L2 outside of the classroom can lead to improved fluency and proficiency in vocabulary and grammar within a short period of time. Therefore, although learning FSs contributed to learners' speaking fluency, in line with previous studies (e.g., McGuire & Larson-Hall, 2017; Taguchi, 2007; Wood, 2009), I believe that there may have been other factors at play that contributed to the control group's improved fluency, beyond the acquisition of collocations alone. Understanding the potential sources of incidental fluency acquisition would contribute valuable

understanding to the complex nature of language learning and the role of implicit language exposure.

Moreover, in order to obtain a more comprehensive understanding of learners' fluency development, future studies could employ longitudinal designs to measure fluency outcomes over an extended period of time. As suggested by Derwing et al. (2006), it is important to consider the limited opportunities for English language practice outside of the classroom, as this factor may have an impact on learners' fluency. By collecting data on learners' language use beyond the instructional setting, researchers can better understand the role of extracurricular language experiences in shaping fluency development. In addition, longitudinal investigations would offer valuable knowledge in terms of the long-term effects of collocation instruction on speaking fluency. Researchers could examine how learners' exposure to and instruction on collocations impact their fluency outcomes over time. This would contribute to our understanding of the effectiveness of collocation instruction in promoting speaking fluency and inform instructional practices in language classrooms.

I believe that addressing these above questions in future research endeavors will contribute to a deeper understanding of the factors influencing speaking fluency and provide valuable implications for instructional practices in SLA. By examining the underlying mechanisms that contribute to fluency gains in both instructed and incidental

learning contexts, language teachers can refine and tailor their pedagogical approaches to optimize language learning outcomes.

CHAPTER 6. Concluding Remarks and Future Directions

This thesis aimed to investigate the impact of explicit instruction on formulaic sequences (FSs), specifically phrasal verbs (PVs) and collocations, in enhancing the writing and speaking fluency of Japanese university learners of English. Through a series of studies, the results have documented the effectiveness of different methods in improving learners' fluency with PVs (Experiment 1-3) and collocations (Experiment 4), employing spaced flashcard-based learning. In particular, the experimental groups that utilized flashcards demonstrated significant advancements in both their writing and speaking fluency, setting them apart from those who did not incorporate this method.

This chapter first provides a concise overview of the key findings from Experiment 1 to 4 and broader implications of this thesis as a cohesive whole. Furthermore, it outlines the limitations inherent in the current thesis and suggests potential avenues for future research.

6.1 Review of Findings and Discussion

Experiment 1 showed the impact of flashcard learning alone, showing significant improvement in learners' PV gap-filling test performance and their ability to retain PVs over the long term. In Experiment 2, the introduction of additional tasks alongside

flashcard-based learning led to a modest yet discernible improvement, highlighting the positive influence of increased time on task. The findings of Experiment 3 revealed the power of combining flashcard learning with speaking practice. This approach resulted in superior PV production in speaking, surpassing the performance of two control groups: one received speaking practice without flashcards, and the other did not receive any form of intervention. Lastly, Experiment 4 provided further evidence for the superior effect of flashcards and speaking practice with collocations in promoting learners' speaking fluency.

Now, let us summarize the implications of this thesis. Initially, throughout Experiment 1-4, this thesis employed spaced retrieval practice using flashcard learning. This approach proves to be a valuable tool for facilitating vocabulary learning and promoting long-term retention. The findings of this study support the notion that spacing out study sessions over time leads to more effective learning outcomes and enhances the ability to retain information in the long run. This method aligns with the spacing effect, which suggests that learning is improved when study sessions are distributed across time rather than condensed into a single session (Kornell, 2009). These findings are consistent with previous studies that have demonstrated the effectiveness of spaced retrieval practice in various domains (e.g., Goossens et al., 2012; Kornell, 2009; Nakata, 2013, 2015).

Although prior research has not specifically explored the application of spaced retrieval practice using flashcards for long-term vocabulary retention, the present study suggests that it can be an effective strategy, even over extended periods of time.

Secondly, Experiment 2 did not find any advantages of incorporating additional activities, including visual aids, alongside flashcards for PV learning. The findings contradicted previous studies that suggested visual assists could be beneficial for vocabulary learning (Bisson et al., 2015; Lin, 2009). However, the total time spent on tasks could be a factor influencing student performance. This indicates that learners' engagement with tasks and the duration of time devoted to them play a significant role in learning PVs. In contrast to these findings, Hill and Laufer (2003) found that the time required to complete the tasks did not have a significant effect on vocabulary learning, although their study focused on form-oriented and meaning-oriented tasks for incidental vocabulary acquisition. Instead, they found that actively engaging with words by exploring their forms, meanings, and relationships on form-oriented tasks led to higher levels of word-related activity, which in turn facilitated vocabulary learning. The conflicting findings in previous research highlight the intricate nature of the relationship between the amount of time spent on a task and its impact on learners' learning outcomes.

However, it is important not to disregard the role of additional activities when compared to simply memorizing word definitions. Effective vocabulary learning requires active engagement in context, including exploring how words are used in different contexts and their relationships with other words (Stahl & Kapinus, 2001). Research conducted by Laufer (2003) supports the effectiveness of activities such as sentence completion, writing words in sentences, and incorporating them into compositions for vocabulary learning. It should be noted that a single session of form-focused vocabulary learning is insufficient for comprehensive learning and long-term retention (Webb et al., 2020). While such a session may result in significant initial gains in understanding the connection between word form and meaning, it should be considered as the initial stage of the word learning process. Continued practice and engagement are necessary to ensure a comprehensive understanding and long-term retention of vocabulary.

The perceived failure of Experiment 2 raises an important question about the worthiness of continuing to explore this aspect of enhancing learning. Although the findings did not support the inclusion of additional activities in the context of flashcard-based PV learning, it is crucial to acknowledge that research is an iterative process.

In light of these outcomes, a decision was made to shift the focus of the thesis towards examining speaking and its potential impact on PV learning. This shift in direction was not an abandonment of the pursuit of effective PV learning methods but rather a response to emerging findings and an opportunity to explore alternative avenues of investigation.

As for the third important finding, which is based on Experiment 3, it highlighted combining memorization through flashcard spaced retrieval practice and output activities, specifically speaking tasks, for effective PV learning. By incorporating flashcards and speaking activities to prompt oral responses, learners could engage in more frequent and active usage of PVs. Engaging in speaking activities provided learners with opportunities to actively use and practice the PVs they had memorized. Supporting this notion, Hulstijn and Laufer (2001) found that the amount of retention was related to the amount of task-induced involvement load. Additionally, engaging actively is crucial in effective vocabulary learning as it involves exploring how words are used in different contexts and understanding their relationships with other words (Stahl & Kapinus, 2001). This aligns with the notion that engaging in speaking activities, as highlighted in this study, allows learners to actively practice and apply PVs into their spoken language use. While there is limited research specifically investigating

the combination of memorization and output activities, collaborative output activities have shown beneficial effects in PV learning (Nassaji & Tian, 2010; Teng, 2017). These activities help learners identify gaps in their language knowledge, directing their attention towards the connection between form and meaning, and receiving valuable feedback from their peers. The findings of this study further reinforce the importance of both memorization and collaborative output activities for effective PV acquisition.

In the final experiment, a finding emerged that consistent and focused use of flashcards could contribute to the development of speaking fluency in a target language, as measured by speech rate (SR) and mean length of run (MLR). By regularly reviewing and engaging with flashcards, learners were able to reinforce their language skills, leading to gradual improvements in their fluency levels. What must also be noted, however, is that even learners who did not receive explicit collocation instruction and specific speaking activities also showed improvement in their fluency. This raises questions about the potential influence of external factors on language acquisition. One plausible explanation for this unexpected outcome could be the learners' exposure to English language input in their environment, such as through university classes or other unidentified sources. Incidental vocabulary acquisition usually requires encountering words multiple times, and studies indicate that it may take more than 20 encounters for

learners to incidentally acquire new vocabulary (Waring & Takaki, 2003). However, given that the participants in this study likely did not encounter the target collocations over 20 times in their EFL environment, it is plausible that other factors contributed to their enhanced fluency. Research has shown that the repetition effect is less effective in listening compared to reading, even with 20 encounters of the target words (Brown et al., 2008). While learning collocations contributed to the learners' speaking fluency, as supported by previous studies (e.g., McGuire & Larson-Hall, 2017; Taguchi, 2007; Wood, 2009), it is likely that other factors also played a role in learner's enhanced fluency, beyond the acquisition of collocations alone. Understanding the underlying factors behind the incidental acquisition of speaking fluency in an English as a Foreign Language (EFL) environment holds great promise in exploring the complexities of language learning and the influence of implicit language exposure.

6.2 Pedagogical Implications

The findings have several pedagogical implications that can inform language instruction and curriculum development:

1. Experiment 1: Using spaced repetition strategies, specifically spaced retrieval with flashcards, in language learning programs can greatly improve how learners learn

and remember vocabulary. By designing study programs that emphasize regular review and recalling of previously learned words, LX learners can strengthen our memory and better retain the vocabulary in the long run (for the term 'LX,' see Footnote 1). Having a learning approach that optimizes the timing of reviewing words is like having a secret formula that helps us remember words more effectively. I believe that incorporating spaced repetition strategies, such as using flashcards, can make a big difference in how learners acquire and remember new words. With a well-designed study program, language educators can encourage students to regularly review and recall the vocabulary they have learned, leading to stronger memory and improved retention over time.

2. Experiment 2: Simply incorporating additional activities, such as those used in the study, alongside flashcard learning may not be sufficient to effectively improve learners' recall and recognition ability of PVs. The total time spent on these activities, rather than the frequency of participation, may lead to significant improvements in student performance. To further investigate the effect of incorporating additional activities alongside flashcard learning, I would explore and examine various factors and instructional strategies for future study. For a future study, for instance, I would try to investigate the ideal amount of time spent on

tasks with the use of technology-based tools that can precisely measure progress. Language teachers could also consider incorporating collaborative learning approaches, where students work together, as this has shown positive effects in previous studies conducted by Nassaji and Tian (2010) and Teng (2017). These studies have already provided evidence supporting the benefits of collaborative or group work in learning PVs compared to individual work. They used specific tasks, such as cloze, editing, and writing tasks, to examine the effects of collaboration on PV learning. However, future studies could expand on this by exploring the optimal amount of time spent on tasks to enhance PV acquisition and long-term retention.

3. Experiment 3: A combination of memorization and speaking activities proved to be an effective approach to enhance the use of PVs. While speaking activities alone can contribute to incidental improvement (van Zeeland & Schmitt, 2013; Vidal, 2011), actively engaging with and memorizing target vocabulary explicitly can be important for successfully incorporating it into language production. Therefore, explicit instruction can play a significant role in consolidating vocabulary knowledge and promoting its effective use in speaking. To improve the effective use of PVs, a recommended approach is to provide explicit instruction with repetition that includes techniques such as the memorization phase used in this

study, where learners were presented with target PVs alongside example sentences and audio to enhance understanding and retention. It is also important to create opportunities for learners to actively participate in speaking activities, allowing them to practice and apply the knowledge they gained during the memorization phase. By adopting this comprehensive approach, language educators could establish a conducive learning environment that maximizes the effective use of PVs by learners.

4. Experiment 4: Incorporating flashcards and speaking activities to learn collocations has been found to have a positive impact on learners' speaking fluency. Personally, as a non-native researcher who actively taught learners using collocations in speaking during this study, I have witnessed first-hand the benefits that these techniques offer in enhancing language fluency. However, as I further examined the study's findings, I realized that there were still aspects that require more in-depth exploration. The varying results observed in the spontaneous speech task indicate that there might be additional factors at play, influencing fluency in different speaking contexts. To foster overall fluency, it might be important to develop a comprehensive speaking curriculum that goes beyond solely collocation instruction. Our goal should be to introduce interactive tasks that encourage

learners to engage actively with the language in authentic ways. For example, incorporating role-plays, discussions, and debates can provide learners with meaningful opportunities to practice their speaking skills. In my study, the control group was from a Debate class, and the class activities might have influenced their improvement in speaking fluency during the spontaneous speech test.

In summary, throughout my research, I explored the incorporation of various strategies and techniques, including spaced repetition strategies, additional instructional strategies, a combination of memorization and speaking activities, and targeted instructional approaches.

6.3 Plans for Further Research

While the findings from the four studies in this thesis provide valuable insights, it is important to acknowledge the limitations of the research that open up avenues for further investigation.

Experiment 1 calls for additional research to address certain limitations related to the format of tests and the use of flashcard learning. In terms of the test format, I found the test format proposed by Fryer et al. (2011) to assess the depth of vocabulary knowledge

quite interesting. They suggest using a format where learners are presented with three sentences and need to select the appropriate word from five choices. This format allows for only one answer which is targeted and can be effective in avoiding ambiguity, especially considering the polysemous nature of PVs. In my study, I used gap-filling tests which required learners to put an appropriate verb and particles with several hints such as the first one or two words of the verb and Japanese translation of the sentence. This could help the learners answer easily. However, incorporating the test format by Fryer et al. (2011) in future studies could potentially enhance the accuracy and validity of the test results.

In addition, future research could explore the acquisition of English PVs among speakers of different language typologies could provide valuable insights into the impact of L1 differences on PV acquisition. Comparative studies involving verb-framed languages (e.g., Japanese, Korean, Turkish, French, Spanish) and satellite-framed languages (e.g., Arabic, Hebrew, Chinese, Greek) would help elucidate the role of typological differences in the ease or difficulty of acquiring PVs. Such investigations could expand of the understanding of the complexities involved in PV acquisition and inform instructional practices for LX learners from diverse linguistic backgrounds.

Additionally, the lack of survey data on participant use of the flashcard learning program between the immediate post-test and the delayed post-test was problematic. It is indeed possible that some learners may have become less motivated to use the flashcards consistently over time, leading to a decline in their memory strength (Brown et al., 2007; Ebbinghaus, 1993). In my study, I addressed this issue by incorporating visual aids in the flashcards, specifically through short animations ('PVgif') for each PV in Experiment 2. These animations were designed to make the learning process more enjoyable and engaging for learners. I believe that this helps them to maintain their motivation for continuous flashcard learning. Another possible effective strategy to enhance motivation is the implementation of systems that allow students to track their progress and receive timely feedback on their flashcard learning. Through the use of digital platforms or applications, learners can log their study sessions and monitor their performance over time. This feature would enable them to visualize their achievements, identify areas for improvement, and receive feedback on their performance. By having a clear understanding of their progress and receiving feedback, students can witness their growth and experience a sense of accomplishment, which in turn would boost their motivation to continue using the flashcards consistently. Additionally, another possible way to increase motivation would be to implement a ranking system that shows learners' total time spent using flashcards on a weekly basis. This ranking system can be

displayed through a leaderboard or a chart that highlights the top performers based on their flashcard usage. By publicly acknowledging and recognizing learners' dedication and effort, a sense of competition can be fostered, motivating learners to invest more time and effort into their flashcard learning. However, it is important to ensure that the ranking system could promote a supportive and non-threatening environment to prevent any negative impact on learners' self-esteem. By incorporating these approaches, it could be possible to create a more motivating and engaging environment for flashcard learning. These strategies could effectively address the issue of declining motivation and help learners maintain their enthusiasm for continuous flashcard practice, leading to improved memory strength and enhanced vocabulary retention.

Raising learners' motivation is an important aspect that has been discussed in Experiment 1 and further emphasized in Experiment 2, which highlights the need for additional investigation into the relationship between additional tasks and PV retention. In order to advance research in this area, it is important to address the limitations identified in the study. Firstly, the limited engagement of students with the activities on *Moodle* raises questions about the potential impact of increased study efforts and active involvement with the materials. Personally, I found the process of creating PVgifs enjoyable, and I believe that this activity contributed to my better understanding and

retention of the PVs. In future studies, it might be beneficial to involve learners in the creation of videos to illustrate PV meanings and encourage them to share their creations with their peers. This would not only enhance learner engagement but also foster a sense of ownership and collaboration in the learning process.

Another limitation that needs to be considered is the voluntary nature of the activities and the low participation rates observed. Most students may not have perceived a necessity to engage with them. The majority of them participated in activities once a week or less, which was inadequate for long-term learning and retention. To address this limitation, future studies could consider making the activities mandatory as part of the class curriculum to assess their impact on PV learning. By making the engagement with the activities a requirement, researchers can evaluate whether the increased participation leads to more significant learning outcomes compared to voluntary involvement. This aspect is closely related to learners' motivation, as the level of commitment and engagement may vary between voluntary and mandatory participation.

The limitations of Experiment 3 raise important considerations for further investigation. One potential limitation of the study could not identify the reason why learners in the groups who received instruction using story descriptions might have been more inclined

to use the target PVs in the descriptive test compared to spontaneous speech. This could be due to their familiarity with the descriptive test format, as they were exposed to similar structured prompts during instruction. As a result, the participants may have felt more comfortable and prepared to use the target PVs in a guided context like the descriptive test. However, in spontaneous speech situations where there were no specific prompts or cues, participants might have been less likely to naturally incorporate the target PVs. This limitation could have implications for the generalizability of the findings to real-life communicative situations where spontaneous language production is required. Future studies could explore ways to encourage the use of target PVs in more spontaneous and unstructured speaking tasks to assess their proficiency in real-time language use.

Another potential limitation is the presence of selection bias in the control groups. The control groups in this study consisted of intact classes, with one group participating in similar speaking activities as the experimental group but without explicit PV instruction, and the other group not engaging in either the activities or PV study. While efforts were made to recruit participants and obtain their consent, it is possible that those who chose to participate were more motivated or had higher proficiency levels, potentially influencing the results. However, it is important to acknowledge the

challenges of implementing random assignment and the associated costs of conducting studies outside of regular class settings. Considering these practical constraints, an alternative approach could be to explore methods that minimize selection bias. For example, researchers could consider using pre-test measures to assess participants' motivation levels and proficiency prior to group assignment. This would allow for more balanced distribution of participants across the control groups, reducing the potential influence of selection bias on the study results.

Lastly, future studies building on the findings of Experiment 4 should consider several aspects to expand knowledge and improve instructional practices related to speaking fluency. One important aspect to consider in future studies is the incorporation of a delayed post-test to measure fluency retention over an extended period. This approach would enable us to capture the long-term effects of collocation instruction on speaking fluency and provide valuable insights into the durability of the acquired skills. By administering a delayed post-test at a later time point, such as several months or even a year after the instruction, researchers can assess the extent to which learners retain and continue to utilize the collocation skills in their speaking.

Furthermore, investigating a wider range of FSs beyond verb-noun and adjective-noun collocations would provide insights into the acquisition of native-like "formulaicity" (Ellis, 2012). This could include lexical bundles, PVs, idioms, and other expressions that shape discourse organization (Boers & Lindstromberg, 2012). Native speakers naturally incorporate all of these FSs into their speech. Would native-like formulaicity be acquired by teaching learners more expressions than only collocations? I believe that each type of FS also contributes to the overall fluency of LX learners. Considering the diverse nature of FSs, it is important to broaden our focus beyond a single type when examining speaking fluency. By examining the influence of different FS types on fluency, researchers can gain a more comprehensive understanding of their role in language production. One potential approach could involve quantifying the occurrence of each FS type in learners' spoken language and analyzing how the frequency of each type correlates with their overall speaking fluency. This quantitative analysis could provide valuable insights into the specific contributions of different FS types to fluency development.

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my primary supervisor, Dr. Jenifer Larson-Hall. Her invaluable advice, continuous support, and patience have been instrumental in guiding me throughout my academic research. Her immense knowledge and plentiful experience have shaped my work and helped me grow as a researcher. I am sincerely appreciative of her unwavering commitment to academic excellence, which has been a constant source of inspiration.

I also extend my gratitude to Dr. Keiko Hirano, my vice-supervisor, for her initial guidance and contributions in shaping the direction of my research.

I am grateful to Dr. Saeko Urushibara for her dedicated time and effort in reading and reviewing my thesis. Her insightful feedback greatly improved the quality and impact of my research.

Furthermore, I am grateful to Mr. Michael McGuire for his assistance in obtaining the data of speech rate and mean length of run used in this thesis. His efforts in providing the necessary data were invaluable and greatly contributed to the completion of my research.

I would like to extend my sincere appreciation to my colleague, Dr. Midori Mitarai. Her consistent motivation has been instrumental in shaping my academic path and has significantly contributed to my personal and professional growth.

My gratitude extends to all the participants who generously participated in my project, as their involvement has been indispensable in making this study possible.

Moreover, I am deeply grateful to Dr. Tetsuji Sato, my dedicated doctor, who has provided continuous medical support throughout my academic path since my undergraduate studies. His exceptional care and treatment have enabled me to overcome the challenges posed by my physical disability, allowing me to pursue and successfully complete my research. I am profoundly thankful for his continuous support, which has played a vital role in my academic achievements.

Lastly, I would like to express my deepest appreciation to my parents, Yukio and Yuko, for their enduring support and encouragement during my PhD studies. Their unwavering belief in me has been a constant source of inspiration.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
ANOVA	Analysis of Variance
AVL	Academic Vocabulary List
BNC	British National Corpus
CI	Confidence Interval
COCA	Corpus of Contemporary American Corpus
EFL	English as a Foreign Language
ESL	English as a Second Language
FDR	False Discovery Rate
FS	Formulaic Sequence
L1	First Language
L2	Second Language
LX	Any foreign language acquired after age 3, to any proficiency level
MLR	Mean Length of Run
PV	Phrasal Verb
PVgif	Short animated clips that illustrate the meaning of a phrasal verb
SLA	Second Language Acquisition
SR	Speech Rate

LIST OF FOOTNOTES

[1] Dewaele (2018, p.3) "suggested using the label 'LX,' meaning any foreign language acquired after the age at which the first language(s) was acquired, that is after the age of 3 years, to any level of proficiency. It is then possible to be either specific and compare the person's L2, L3, or L4, or to make a more global statement about the person's LXs." I would also like to refer to a language that learners study as LX.

[2] The department of the university has a *hensachi* (deviation score) ranking between 50.0-52.5, which corresponds to a Common European Framework of Reference for Languages (CEFR) level between B1-B2.

[3] *Quizlet Plus* is a paid application. *Quizlet* has since discontinued their spaced repetition program. This feature was only available if students used a computer to review words and typed in the answers, so not all students might have availed themselves of this feature since studying vocabulary on their mobile devices was also possible. One program which does provide spaced repetition with mobile devices is *Anki*, but it takes a little more work than *Quizlet* to install and set up. In addition, technical burden that would reduce the risk of student participation could be considered, so I chose the *Quizlet* system.

[4] *Moodle* is "a learning platform designed to provide educators, administrators and learners with a single robust, secure and integrated system to create personalized learning environments (Moodle, 2020)". It is known as one of the e-learning systems, which is the Learning Management System (LMS).

[5] One-Way Analysis of Variance: "this is used when you want to test whether the scores of three or more groups or levels differ statistically" (Larson-Hall, 2015, p. 180).

[6] Note that although the correlation coefficients are similar, this is a different

correlation than the one above between pre-test and post-test score.

[7] University A and University B were estimated to have a Common European Framework of Reference for Languages (CEFR) level between B1-B2 and B1-C1, respectively, based on their *hensachi* (deviation score) calculated from the average score on the standardized Center Test (now known as the Unified University Entrance Examination). It is worth noting that the test of English subject only assesses reading and listening skills.

[8] *Work out* has more than two meaning. For example, 1) Plan, devise or think about STH carefully or in detail, 2) Exercise in order to improve health or strength (Garnier & Schmitt, 2014, in PHaVE List, p. 7). The present study used the first meaning in the list. However, some participants used *work out* as the second meaning in the second story. In this case, the PV was not counted because the meaning was not a targeted one.

[9] In terms of SR and MLR, which take syllable count into account, it is important to consider the potential difference between collocations and lexical items. For instance, collocations like 'senior citizens' are likely to have more syllables compared to lexical items like 'the elderly,' which could impact the analysis. In this study, the analysis of SR and MLR aimed to investigate the impact of incorporating a speaking component with collocations on learners' speaking fluency. By examining these temporal variables, the study aimed to gain insights into how the intervention influenced the pace and flow of learners' speech.

REFERENCES

Acar, A., & Kayaoglu, M. N. (2020). MOODLE as a potential tool for language education under the shadow of COVID-19. *Eurasian Journal of Educational Research*, *90*, 67-82. https://doi.org/10.14689/ejer.2020.90.4

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*(7), 861-876. http://dx.doi.org/10.1002/acp.1391

Akbary, M., Shahriari, H., & Fatemi, A. H. (2016). The value of song lyrics for teaching and learning English phrasal verbs: A corpus investigation of four music genres. *Innovation in Language Learning and Teaching*, *12*(4), 344-356. https://doi.org/10.1080/17501229.2016.1216121

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

Alangari, M., Jaworska, S., & Laws, J. (2019). Who's afraid of phrasal verbs? The use of phrasal verbs in expert academic writing in the discipline of linguistics. *Journal of English for Academic Purpose*s, *43*, 100814. https://doi.org/10.1016/j.jeap.2019.100814

ALC PRESS INC. (n.d.). Example sentences of phrasal verbs. In *Eijiro on the web*.

https://eow.alc.co.jp/

Anderson, C. J., Stewart, J., & Fryer, L. (2009). Coordinating a curriculum-wide vocabulary program. In *Proceedings of the 17th Annual KOTESOL International Conference* (pp. 261-268). Seoul, Korea: Korea Teachers of English to Speakers of Other Languages (Korea TESOL / KOTESOL).

Baddeley, A. D. (1997). *Human memory: Theory and practice*. (Revised Edition). East Sussex, UK: Psychology Press.

Baicheng, Z. (2009). Do example sentences work in direct vocabulary learning? *Issues in Educational Research*, *19*(2), 175-189.

Bialystok, E. (1983). Inferencing: Testing the "Hypothesis-Testing" hypothesis. In H. W. Seliger & M. H. Long (Eds.), *Classroom-oriented research in second-language acquisition* (pp. 104-123). Rowley, MA: Newbury House.

Biber, D. (2009). A corpus driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology, 68*(7), 1306-1326. https://doi.org/10.1080/17470218.2014.979211

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, H. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*(3), 245-261. https://doi.org/10.1191/1362168806lr195oa

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83-110. https://doi.org/10.1017/S0267190512000050

Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539-576. https://doi.org/10.1037/0033-295x.114.3.539

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136-163.

Cadierno, T. (2008). Learning to talk about motion in a foreign language. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 239-275). New York, NY: Routledge.

Cambridge University Press. (n.d.). Example sentences of phrasal verbs. In *Cambridge dictionary*. https://dictionary.cambridge.org/

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL*

teacher's course (2nd ed.). Boston, MA: Heinle and Heinle.

Cepeda, N. J., Pashle, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380. https://doi.org/10.1037/0033-2909.132.3.354

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, *25*(3), 301-315. https://doi.org/10.1016/S0346-251X(97)00024-9

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45-61. https://doi.org/10.1017/S0267190512000074

Costa, C., Alvelos, H., & Teixeira, L. (2012). The Use of Moodle E-Learning Platform: A Study in a Portuguese University. *Procedia Technology*, *5*, 334-343. https://doi.org/10.1016/j.protcy.2012.09.037

Crandell, E. R. (2017). Quizlet flashcards for the first 500 words of the academic vocabulary list. *All Theses and Dissertations*. 6335. https://scholarsarchive.byu.edu/etd/6335

Crothers, E., & Suppes, P. (1967). *Experiments in second language learning*. New York, NY: Academic Press.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215-235.

https://psycnet.apa.org/doi/10.1002/(SICI)1099-0720(200005/06)14:3%3C215::AID-ACP640%3E3.0.CO;2-1

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition*, *7*(1), 73-79. https://doi.org/10.1017/S0272263100005167

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Davies, M., & Gardner, D. (2013). *The academic vocabulary list.* Brigham Young University.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, *34*(2), 183-193. https://doi.org/10.1016/j.system.2006.01.005

Dewaele, J.-M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker.' *Applied Linguistics, 39*(2), 236-240. https://doi.org/10.1093/applin/amw055

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163-188. https://doi.org/10.1177/0267658309349431

Ebbinghaus, H. (1993). *Memory: A contribution of experiential psychology*. (H. A.

Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University.

Ellis, N. C. (2012). Formulaic language in second language acquisition: Zipf and phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44. https://doi.org/10.1017/S0267190512000025

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*(2), 285-301. https://doi.org/10.1017/S0272263199002077

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Text*, *20*(1), 20-62. https://doi.org/10.1515/text.1.2000.20.1.29

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal, 36*(2), 239-248. https://doi.org/10.1080/09571730802390759

Fryer, L. K., Stewart, J., Anderson, C. J., Bovee, H. N., & Gibson, A. (2011). Coordinating a vocabulary curriculum: Exploration, pilot, trial and future

directions. *JALT2010 Conference Proceedings*. Tokyo: JALT. https://jalt-publications.org/files/pdf-article/jalt2010proc-10.pdf

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*(2), 339-359. https://doi.org/10.1002/j.1545-7249.2007.tb00062.x

Garnier, M., & Schmitt, N. (2014). The phase list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, *19*(6), 645-666. https://doi.org/10.1177/1362168814559798

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29-44. https://doi.org/10.1016/j.system.2016.04.004

Golding, J. M., Wasarhaley, N. E., & Fletcher, B. (2012). The use of flashcards in an introduction to psychology class. *Teaching of Psychology*, *39*(3), 199-202. https://doi.org/10.1177/0098628312450436

Goossens, N. A. M. C., Camp, G., & Verkoeijen, P. P. J. L. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology*, *24*(8), 965-971. https://doi.org/10.1080/20445911.2012.722617

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: theory, analysis, and applications* (pp.

79-100). Oxford: Oxford University Press.

Groot, P. J. M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, *4*(1), 60-81.

Gyllstad H., & Schmitt, N. (2019). Testing Formulaic Language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 174-191). London, UK: Routledge. http://dx.doi.org/10.4324/9781315206615-10

HarperCollins Publishers Limited. (n.d.). Example sentences of phrasal verbs. In *Collins English dictionary*. https://www.collinsdictionary.com/dictionary/english

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

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review / La revue canadienne des langues vivantes, 61*(3), 355-382. https://doi.org/10.3138/cmlr.61.3.355

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge

University Press. http://dx.doi.org/10.1017/CBO9781139524780.011

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539-558. https://doi.org/10.1111/0023-8333.00164

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133-166). Hillsdale, NJ: Lawrence Erlbaum Associates.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research*, *30*(1), 138-149. https://doi.org/10.1086/374692

Jorgensen, J. C. (1990). The psychological reality of word senses. *Journal of Psycholinguistic Research*, *19*(3), 167-190. https://doi.org/10.1007/BF01077415

Kanayama, K., & Kasahara, K. (2016). The effects of expanding and equally-spaced retrieval practice on long-term L2 vocabulary retention. *Annual Review of English*

Language Education in Japan, *27*, 217-232. https://doi.org/10.20581/arele.27.0_217

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, *21*(6), 1544-1550. http://dx.doi.org/10.3758/s13423-014-0636-z

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1250-1257. https://doi.org/10.1037/a0023436

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772-775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 704-719. https://doi.org/10.1037/0278-7393.33.4.704

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*(3), 365-386. https://doi.org/10.1177/1362168808089922

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Learning Language*, *58*(2), 285-325. https://doi.org/10.1111/j.1467-9922.2008.00442.x

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145-164. https://doi.org/10.1016/j.system.2004.01.001

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297-1317. https://doi.org/10.1002/acp.1537

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219-224. https://doi.org/10.3758/BF03194055

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585-592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness*, *3*(2), 73-93. https://doi.org/10.1080/09658416.1994.9959845

Kuiken, F., & Vedder, I. (2002). The effect of interaction in acquiring the grammar of a

second language. *International Journal of Educational Research, 37*(3-4), 343-358. http://dx.doi.org/10.1016/S0883-0355(03)00009-0

Kurtyka, A. (2001). Teaching English phrasal verbs: A cognitive approach. In M. Pütz & R. Dirven (Eds.), *Language pedagogy* (pp. 31-54). Berlin: De Gruyter.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R.* New York: Routledge.

Larson-Hall, J. (2015). Bootstrapped and means-trimmed one-way ANOVA and multiple comparisons in R. In *A guide to doing statistics in second language research using SPSS and R* (p. 331). https://s3-eu-west-1.amazonaws.com/s3-euw1-ap-pe-ws4-cws-documents.ri-prod/9781138024571/chapters/ch9/1_Bootstrapped_One-Way_ANOVA_in_R.JLH.pdf

Larson-Hall, J., & Herrington, R. (2010). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics*, *31*(3), 368-390. https://doi.org/10.1093/applin/amp038

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567-587. http://dx.doi.org/10.3138/cmlr.59.4.567

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

Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 163-176). Abingdon, Oxon.: Routledge.

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RECL Journal*, *28*(1), 89-108. https://doi.org/10.1177/003368829702800106

Leeser, M. J. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research*, *8*(1), 55-81. https://doi.org/10.1191/1362168804lr134oa

Le-Thi, D., Rodgers, M. P. H., & Pellicer-Sánchez, A. (2017). Teaching formulaic sequences in an English- language class: The effects of explicit instruction versus coursebook instruction. *TESL Canada Journal*, *34*(3), 111-139. https://doi.org/10.18806/tesl.v34i3.1276

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85-102. https://doi.org/10.1016/j.jslw.2009.02.001

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, *54*(2) 193-226. https://doi.org/10.1111/j.1467-9922.2004.00254.x

Lin, C. (2009). Learning action verbs with animation. *The JALT CALL Journal, 5*(3), 23-40. https://doi.org/10.29140/jaltcall.v5n3.84

Littlemore, J. (2003). The effect of cultural background on metaphor interpretation. *Metaphor and Symbol*, *18*(4), 273-288. https://doi.org/10.1207/S15327868MS1804_4

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly*, *45*(4), 661-688. https://doi.org/10.5054/tq.2011.247707

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition*, *15*(3), 257-280. https://doi.org/10.1080/13825580701322171

Lu, M., Farhat, J. H., & Beck Dallaghan, G. L. (2021). Enhanced learning and retention of medical knowledge using the mobile flash card application Anki. *Medical Science Educator*, *31*(6), 1975-1981. https://doi.org/10.1007/s40670-021-01386-9

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, *27*(3), 405-430. https://doi.org/10.1093/applin/ami051

Mair, P., & Wilcox, R. (2020). Robust Statistical Methods in R Using the WRS2 Package. *Behavior Research Methods*, *52*, 464-488. https://doi.org/10.3758/s13428-019-01246-w

McDonough, K., & Mackey, A. (2000). Communicative tasks, conversational interaction and linguistic form: An empirical study of Thai. *Foreign Language Annals*, *33*(1), 82-92. https://doi.org/10.1111/j.1944-9720.2000.tb00893.x

McGuire, M., & Larson-Hall, J. (2017). Teaching formulaic sequences in the classroom: Effect on spoken fluency. *TESL Canada Journal*, *34*(3), 1-25. https://doi.org/10.18806/tesl.v34i3.1271

Mizumoto, A. (2015). Langtest (Version 1.0) [Web application]. http://langtest.jp

Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40-63). Cambridge: Cambridge University Press.

Munoz, A., & Mackay, J. (2019). An online testing design choice typology towards

cheating threat minimisation. *Journal of University Teaching & Learning Practice, 16*(3). https://doi.org/10.53761/1.16.3.5

Nakata. T. (2008). English vocabulary learning with word lists, word cards, and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL*, *20*(1), 3-20. https://doi.org/10.1017/S0958344008000219

Nakata, T. (2013). Optimising second language vocabulary learning from flashcards (Unpublished doctoral dissertation). Victoria University of Wellington, New Zealand. https://www.academia.edu/4557533

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, *37*(4), 677-711. https://doi.org/10.1017/S0272263114000825

Nakata, T. (2017). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition*, *39*(4), 653-679. https://doi.org/10.1017/S0272263116000280

Namaziandost, E., Sawalmeh, M. H. M., & Soltanabadi, M. I. (2020). The effects of spaced versus massed distribution instruction on EFL learners' vocabulary recall and retention. *Cogent Education*, *7*(1), 1792261.

https://doi.org/10.1080/2331186X.2020.1792261

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

Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: a review of the research. *RELC Journal*, *13*(1), 14-36. https://doi.org/10.1177/003368828201300102

Nation, I. S. P. (1990). *Teaching and learning vocabular*y. New York, NY: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*(2), 355-380.

Ng, C. F., & Ng, P. K. (2015). A review of intrinsic and extrinsic motivations of ESL learners. *International Journal of Languages, Literature and Linguistics, 1*(2), 98-105. https://doi.org/10.7763/IJLLL.2015.V1.20

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. https://doi.org/10.1177/1362168816639619

Ortega, L., & Brynes, H. (2008). *The longitudinal study of advanced L2 capacities: An introduction*. New York, NY: Routledge.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26-45. https://doi.org/10.1017/S0267190505000024

Oxford, R., & Crookall, D. (1990). Vocabulary learning: a critical analysis of techniques. *TESL Canada Journal*, *7*(2), 9-30. https://doi.org/10.18806/tesl.v7i2.566

Oxford University Press. (n.d.). Example sentences of phrasal verbs. In *Oxford learner's dictionaries*. https://www.oxfordlearnersdictionaries.com/

Paivio, A. (1990). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-225). New York: Longman.

Pearson Education Limited. (n.d.). Head. In *Longman dictionary of contemporary English*. https://www.ldoceonline.com/jp/dictionary/head

Pearson Education Limited. (n.d.). Example sentences of phrasal verbs. In *Longman*

dictionary of contemporary English. https://www.ldoceonline.com/

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*(8), 1917-1927. https://doi.org/10.3758/BF03192925

Rieber, L. P. (1989). The effects of computer animated elaboration strategies and practice on factual and application learning in an elementary science lesson. *Journal of Educational Computing Research*, *5*(4), 431-444. https://doi.org/10.2190/961X-QQEQ-7PVK-UK9E

Rudzka-Ostyn, B. (2003). *Word power: Phrasal verbs and compounds*. Berlin: De Gruyter.

Saito, K., & Hanzawa, K. (2016). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research*, *22*(4). https://doi.org/10.1177/1362168816679030

San-Mateo-Valdehíta, A., & Chacón-García, C. (2019). Learning word class in a second language through vocabulary learning activities: definition-choosing, gap-filling, and sentence-writing. *Journal of Spanish Language Teaching*, *6*(1), 49-63. https://doi.org/10.1080/23247797.2019.1590001

Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: internal structure and external connections. In R. L. Oxford (Ed.), *Language learning*

motivation: Pathways to the new century (pp. 14-87). Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. https://doi.org/10.1177/1362168808089921

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

Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. D. Cock, G. Gilquin, & M. Paquot (Eds.), *A test of corpora: In honour of Sylviane Granger* (pp. 173-208). John Benjamins Publishing Company. https://doi.org/10.1075/scl.45.12sch

Shiiba, A. (2021). To what extent does meaning prototypicality of the particle of phrasal verbs affect learning? *JALT Vocabulary Education and Research Bulletin*, *10*(2), 2-16. https://jaltvocab.weebly.com/uploads/3/3/4/0/3340830/verb-vol10.2.pdf

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 907-912. https://doi.org/10.1037/0278-7393.27.4.907

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Slobin, D. I. (2005). Linguistic representations in motion events: What is signifier and what is signified? In C. Maeder, O. Fischer, & W. J. Herlofsky (Eds.), *Iconicity in language and literature 4: Outside-in - inside-out* (pp. 307-322). Amsterdam: Benjamins.

Stahl, S. A., & Kapinus, B. (2001). *Word power: What every educator needs to know about teaching vocabulary*. Washington, D.C.: National Education Association.

Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does 'chunking' foster chunk-uptake? In De Knop, S., Boers, F., & De Rycker, A. (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 99-117). Berlin: De Gruyter Mouton.

Stoeckel, T., Stewart, J., McLean, S., Ishii, T., Kramer, B., & Matsumoto, Y. (2019). The relationship of four variants of the Vocabulary Size Test to a criterion measure of meaning recall vocabulary knowledge. *System*, *87*, 102161. https://doi.org/10.1016/j.system.2019.102161

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*, 244-253. https://doi.org/10.3758/MC.38.2.244

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.),

Input in second language acquisition (pp. 235-253). Rowley, MA: Newbury House.

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research, 11*(4) 433-457. https://doi.org/10.1177/1362168807080962

Talmy, L. (2000). *Toward a cognitive semantics, volume II: Typology and process in concept structuring*. Cambridge, MA: The MIT Press.

Taniwaki, Y., & Tono, T. (2009). Kudōshi [Phrasal verbs]. In T. Kageyama (Ed.), *Nichiei taisho: Keiyōshi, fukushi no imi to kōbun* [English-Japanese contrasts: Semantics and syntax of adjectives and adverbs] (pp. 293-324). Tokyo, Japan: Taishūkan Shoten.

Teng, M. F. (2017). The effectiveness of group, pair and individual output tasks on learning phrasal verbs. *The Language Learning Journal*, *48*(2), 187-200. https://doi.org/10.1080/09571736.2017.1373841

Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning and cognition*. Cambridge University Press: Cambridge.

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

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

Waring, R. & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130-161. https://nflrc.hawaii.edu/rfl/item/78

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *19*(3), 287-307.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33-52. https://doi.org/10.1017/S0272263105050023

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63-81. https://doi.org/10.1177/1362168806072463

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232-245. https://nflrc.hawaii.edu/rfl/item/178

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal*, *104*(4), 715-738. https://doi.org/10.1111/modl.12671

Weisman, S. R. (1992, April 27). How do Japan's students do it? They cram. *The New York Times*. https://www.nytimes.com/1992/04/27/world/how-do-japan-s-students-do-it-they-cram.html

White, B. J. (2012). A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal*, *96*(3), 419-438. https://doi.org/10.1111/j.1540-4781.2012.01365.x

Wong, J., & Waring, H. Z. (2010). *Conversation analysis and second language pedagogy: A guide for ESL/EFL teacher*s. New York: Routledge.

Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT Journal*, *29*(2), 209-230. https://doi.org/10.37546/JALTJJ29.2-3

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, *12*(1), 39-57.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice, *Applied Linguistics*, *21*(4), 463-489. https://doi.org/10.1093/applin/21.4.463

Wright, A. (1989). *Pictures for language learning*. Cambridge: Cambridge University Press.

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *TESOL Quarterly*, *44*(2), 250-273. http://dx.doi.org/10.2307/27896724

PHaVE Appendix A

This appendix contains sentences and their corresponding Japanese translations for 100 phrasal verbs used in the pre-tests and post-tests of Experiment 1. Typically, each phrasal verb is accompanied by a single definition representing the most common meaning sourced from the PHaVE List (Garnier & Schmitt, 2014). However, in instances where there was a lack of ample examples for the most common definitions or when the first and second definitions significantly overlapped, both definitions were provided. The number assigned to each phrasal verb indicates its rank in the PHaVE List.

Note: STH refers to “something,” and SB refers to “somebody.”

First semester

Phrasal verb and its meaning	Sentences
1. go on (in set #1) Happen, take place	What the hell is going on? 一体何が起こっているの？
34. look back (in set #1) Think of STH again; reconsider STH past	We looked back on her life in remembrance. 私たちは、彼女の思い出とともに、彼女の人生を振り返った。
63. pull up (in set #1) Stop or cause a vehicle to stop	The truck is pulling up to load the cargo now. そのトラックは今、積荷するために停まっている。
103. run out of (in set #1) Use STH (or become used) completely so that nothing is left	I didn’t expect to run out of time so quickly. こんなに早く時間を使い切ってしまうなんて予想していなかった。
2. pick up (in set #2)	The volunteer groups pick up trash in parks every week.

Get or take SB/STH from a place	そのボランティア団体は、毎週公園でゴミ拾いをしている。
37. take over (in set #2) Gain control, management, or possession of STH/SB	They're fighting against the rebel group that is trying to take over the country. 彼らは、国を乗っ取ろうとしている反乱軍と戦っている。
61. cut off (in set #2) Remove a part of STH by cutting it	Cutting off the top of the tree can harm it rather help it. その木のてっぺんを切り取ってしまうと、木を助けるどころか傷つけてしまうよ。
94. turn down (in set #2) Refuse or dismiss	She turned down the invitation to today's business drinking party. 彼女は、今日行われる会社の飲み会へのお誘いを断った。
11. set up (in set #3) Establish or create STH; arrange for STH to happen or exist	The fund was set up to compensate for those who suffered financial losses. その基金は金銭的損失を被った人たちへの補償として設けられた。
42. look down (in set #3) Lower one's eyes to see what is below	No one is speaking, just looking down at the floor. 誰も喋っておらず、ただ床を見下ろしている。
89. come over (in set #3) Come to a place or area, esp. towards SB or to join SB	She came over to his table and sat next to him. 彼女は彼のテーブルへとやって来て、彼の隣に座った。
113. play out (in set #3) Happen or develop; be enacted or performed	His life story was played out in the show. 彼の人生の物語がそのショーで展開された。
15. take on (in set #4) Undertake or handle	The number of older Americans taking on student debt on behalf of their children and grandchildren has quadrupled in the past decade. 子供や孫の代わりに学資金を受け持っているアメリカ人高齢者の数は、過去 10 年で 4 倍になっている。
38. hold up (in set #4) Hold STH in a high position	On the counter, the glass looks clean, but holding up the glass in the air, it doesn't look so clean. カウンターではグラスは綺麗に見えるが、持ち上げてみると、そんなに綺麗ではない。
95. back up (in set #4) Move or drive backwards a short way	He backed up a short way to park his car in an appropriate place. 彼は適切な場所に駐車するため少し後ろに下がった。
111. stand out (in set #4) Distinguish oneself/itself by being better, more significant or more impressive than other people/things	She is wearing false eyelashes to make her eyes stand out more. 彼女は目をさらに際立たせるために、つけまつ毛を付けている。

19. get back (in set #5) Return to a place, position, state, activity, conversation topic	It will be very hard to get back into the game after you are injured. 怪我をしたら、試合に戻るのがとても難しくなるよ。
48. open up (in set #5) Make STH become available or possible, less limited	A whole new world of music opened up my vision of the future. この全く新しい音楽の世界が、私自身の将来像を広げてくれた。
67. turn over (in set #5) Surrender possession or control to SB/STH	Investigators have turned over evidence to the FBI for technical assistance. 捜査員は、技術的な援助のために FBI に証拠を渡した。
117. get through (in set #5) Succeed in reaching a physical destination or stage	It took time to get through the maze of mirrors in the amusement park. 遊園地の鏡の迷路を抜け出すのに時間がかかった。
21. figure out (in set #6) Come to understand or determine STH	The teacher-librarian serves as a guide to help students figure out the answer on their own. その司書教諭は、生徒たち自身で答えを見つけ出せるように導く仕事をしている。
46. look out (in set #6) Look outside, or at the horizon	The posters feature a woman looked out the back window of a van. そのポスターは、バンの後ろ窓から外を見る女性が特徴的である。
71. line up (in set #6) Form or make SB/STH form into a line	The kids always line up to wash their hands before dinner. 子供たちはいつも、夕食前に手を洗うため並んでいる。
91. start out (in set #6) Start a life, existence, profession, or course of action in a particular way or by doing a particular thing	In this book, the story of “Star Wars” starts out on the planet of Naboo, which has been blockaded by the Trade Federation. この本では、「スター・ウォーズ」の物語は、貿易連合によって封鎖されていたナブーの惑星で始まる。
118. hold back (in set #6) Decide not to do or say STH	He had to hold back his anger at the president’s stupid comments. 彼は大統領のばかばかしいコメントに対する怒りを抑えなければならなかった。
22. sit down (in set #7) Move from a standing position to a sitting position	She sat down and spread the newspaper out over the breakfast table. 彼女は朝食の卓で新聞を広げて座った。
36. carry out (in set #7) Perform or complete	American troops will carry out the mission. アメリカ軍はその任務を遂行するつもりだ。
53. be/get caught up with (in set #7) Become involved in STH which prevents SB from making progress or moving forward	She shed many tears for the people caught up in this horrendous incident. 彼女は、この恐ろしい事故に巻き込まれた人たちのために涙をたくさん流した。

97. send out (in set #7) Mail, send or distribute to a number of people	Hughes apologized for his comment, via an e-mail sent out last night. ヒューズは昨夜送ったメールを通して、自分のコメントについて謝った。
10. grow up (in set #8) Gradually advance in age and maturity	In the generation I grew up in, snowboarding was really popular. 私が育った世代では、スノーボードがとても人気だった。
29. work out (in set #8) Plan, devise or think about STH carefully or in detail	They should be able to work out their differences. 彼らは、お互いによく考えて折り合いをつけることができるはずだ。
88. get down to (in set #8) Begin to pay serious attention to STH	I don't want to be vague anymore; let's get down to the specifics on this agreement. これ以上あいまいにしたくない。この協定についての詳細へ真剣に取り組みましょう。
109. lay down (in set #8) Put STH away or down on a surface, esp. because one has stopped using it	Silvia surrendered and laid down her weapons. シルビアは今降伏し、武器を置いた。
25. come on (in set #9) Said to encourage SB to try harder, or do or say STH	Oh, come on, Mr. Smith. Her words must be just a joke. 大丈夫だよ、スミスさん。彼女の言葉はただのジョークなはずですよ。
51. put out (in set #9) Make STH known or accessible to the public	Local authorities put out a call for blood donations. 地方自治体は、献血が必要であることを発表した。
78. pay off (in set #9) Pay the complete amount of STH	She will not be able to pay off her credit card bill in full this month. 彼女は今月のクレジットカードの最低返済料金を完済できない。
108. step back (in set #9) Move back by lifting one's foot and putting it down backwards	He shook her hand and then stepped back. 彼は彼女と握手をして、一歩後ろに下がった。
17. make up (in set #10) Form the whole of an amount or entity	Women make up less than 10% of the students in this technical college. 女性はこの高等専門学校では 10%もいない。
32. go ahead (in set #10) Proceed with a course of action without further hesitation	The Nursing and Midwifery Council (NMC) has gone ahead with its plan to raise its annual registration fee in spite of stiff opposition. 看護師助産師会は、猛反対にもかかわらず、年間登録料を引き上げる計画を進めた。
62. turn back (in set #10) 1) Turn around so as to face the opposite direction 2) Go back (or make SB/STH	You can't turn back the tide of history. 歴史の流れは戻せないよ。

go back) in the direction SB/STH has come from	
86. hang out (in set #10) Spend time relaxing or enjoying oneself	I found it to be a nice, peaceful place to hang out. People play roller hockey there. リラックスして時間を過ごすのに、落ち着いた場所を見つけた。そこではローラーホッケーをしている人たちがいるよ。
112. come along (in set #10) Appear or arrive; come into existence	Once children come along a family needs more income. 一度子供ができると、家族は収入がさらに必要になる。
12. turn out (in set #11) Prove or be discovered to happen or be	The world has turned out to be a far more complicated place than we imagined. 世界は私たちが想像していたよりも、かなり複雑な場所だということが分かった。
57. keep up (in set #11) Move, progress or increase at the same rate or pace as SB/STH	I made an effort to keep up with popular culture. 日本のポップカルチャーに遅れをとらないよう努力した。
74. go over (in set #11) Move towards a place or person, esp. by crossing an area	I will jump up and go over this wall to take a short cut. 俺は近道するためにこの壁を飛び越えるよ。
107. bring about (in set #11) Cause to happen or emerge, esp. STH positive	The Industrial Revolution brought about increased use of iron and other metals. 産業革命は、鉄やほかの資材の使用増加をもたらした。
13. get out (in set #12) Leave a container or make SB/STH leave a container	I want to take a week off and get out of this tiny town next month. 来月 1 週間の休暇を取って、この狭い町を出たいな。
41. take up (in set #12) Use a particular amount of space, time or effort	Four team members took up positions at the corners of the blue mat. 青いマットの端で 4 人のチームメンバーが場所を陣取っていた。
59. reach out (in set #12) Stretch an arm in order to hold, touch, or get STH that is within short distance	Her fingers reached out to touch my shoulder. 彼女は手を伸ばして、指が私の肩に触れた。
80. break up (in set #12) End or cause to end or fail	Someone called the police and they came and broke up the rowdy party. 誰かが警察を呼んだので、警察が来てその騒々しいパーティを止めさせた。

Second semester

Phrasal verb and its meaning	Sentences
47. bring in (in set #13) Bring STH to a place or situation	A good job will bring in money so a family can live at a reasonable standard. 良い仕事はお金をもたらし、適切な生活基準で生活することができる。
70. turn up (in set #13) Yield; be (or make STH be) found, discovered, or noticed	Our special investigation turned up some surprising results. 我々の特別捜査は、いくつかの驚くべき結果を発見した。
75. hang up (in set #13) Finish a conversation on the telephone by putting the receiver down or switching the phone off	Megan hung up the telephone with a tired sigh. メーガンは疲れたようなため息をついて、電話を切った。
119. write down (in set #13) Record information on paper	Writing down your thoughts may be as effective as saying them out loud. 自分の考えを書き留めるのは、声に出すのと同じくらい効果的だと思うよ。
27. show up (in set #14) Make an appearance at a social or professional gathering	I got very excited when the celebrities showed up. 有名人が現れた時には、とてもワクワクした。
64. set out (in set #14) Start doing or working on STH, esp. with a particular goal in mind	The team set out to design its own specialized camera. そのチームは独自の特殊なカメラのデザインに乗り出した。
92. call out (in set #14) Speak or utter loudly	A feeble voice called out. “Help me!” 弱々しい声で叫んだ「助けて！」
116. walk out (in set #14) Leave a place or event, especially suddenly or angrily	I grabbed my backpack from under my bed, made sure I had keys and wallet, and walked out the door. ベッドの下から荷物を大急ぎで取って、鍵と財布を入れたのを確認し、ドアから出て行った。
28. take off (in set #15) Remove STH (esp. piece of clothing or jewelry from one’s body)	People in other countries don’t take off their shoes even when going to sleep. 別の国では、寝る時でさえも靴を脱がない。
52. look around (in set #15) Examine a place or one’s surroundings so as to view	I looked around a place to hide. 私は隠れために場所を見回した。

what it might contain or look for a particular thing	
73. lay out (in set #15) Describe or explain STH clearly or in detail, esp. officially and in writing	The document clearly lays out nine specific proposed amendments to the law. この文書は、その法律に対する 9 つの具体的な修正案を明確に提示している。
114. break out (in set #15) Start suddenly, esp. STH undesirable and unpleasant	Firefighters are still searching for victims; many may have been asleep when the fire broke out. 消防士は未だ、被災者を探している。おそらく多くの人は、火事が発生した時には眠っていただろう。
9. point out (in set #16) Direct attention toward STH (fact, idea, information)	They're pointing out the fact that politicians are hypocrites. 彼らは、政治家が偽善者であるという事実を指摘している。
30. stand up (in set #16) Rise to a standing position after sitting or lying down	The audience stood up and applauded to the singer. 観客は立ち上がって、その歌手に拍手を送った。
56. get off (in set #16) Go away from, leave (train, bus, aircraft, lift)	We had arrived in Houston and I got a call as I got off the plane. ヒューストンに到着し、飛行機を降りると電話がかかってきた。
79. hold out (in set #16) Move one's hand or an object in one's hand forward or towards SB, in order to grab or give STH	Jane held out her hand and said. "Nice to meet you." ジェーンは自分の手を伸ばして、「初めまして」と挨拶した。
105. shut up (in set #16) Stop (or make SB/STH stop) talking or making a noise	I hate when the people can't shut up about the stupid tabloid headlines. くだらないタブロイド紙の見出しについて、黙っていられないのが、私は嫌いだ。
16. give up (in set #17) Stop doing or having STH; abandon (activity, belief, possession)	Politicians don't usually give up their power willingly. 政治家というものは、快く自分の権力を手放したりはしない。
26. go down (in set #17) Move down to a lower level or position	The damaged plane went down into the water off the coast. その損傷した飛行機は、沿岸沖の水上に降りた。
55. break down (in set #17) Stop working or functioning; fail or collapse (vehicle, device, relationship, negotiations)	If teachers really want to break down barriers with their teenage students, they might try using popular music to engage them. 先生が 10 代の生徒たちとの壁を本当に壊したければ、彼らを引き込むためにポピュラー音楽を使おうとするだろう。
77. hold on (in set #17) Refuse to let go of STH	How long do people hold on to their current iPhones? みんな、今の iPhone をどのくらい持ち続けているの？

106. turn off (in set #17) Stop a piece of equipment working temporarily or a supply flowing by turning a tap, pressing a button, or moving a switch	Do you have a system that can turn off your engine remotely if the car's stolen? もし車が盗まれた場合に、遠隔でエンジンを止められるシステムなんてあるの？
4. come up with (in set #18) Bring forth or produce	Many people are working hard to come up with solutions to a complex problem. 多くの人は、複雑な問題への解決策を考え出そうと頑張っている。
33. go up (in set #18) Become higher in value; increase	Food prices went up to compensate for higher fuel cost. 高い燃料費を埋め合わせるために、食品価格が上がった。
68. slow down (in set #18) Move, proceed or progress at a slower pace (vehicle, economy)	The world economy is slowing down due to uncertainties in Europe and US markets. ヨーロッパとアメリカ市場における不安定のために、世界経済は低迷している。
81. bring out (in set #18) Make a particular detail, quality or feeling more noticeable than it usually is	A touch of cinnamon brings out the sweetness of apple pies. 少しのシナモンだが、アップルパイの甘味を引き立ててくれる。
35. wake up (in set #19) Become (or make SB become) conscious again after being asleep	He always goes to bed early and wakes up before dawn. 彼はいつも早く就寝し、夜明け前に目覚める。
69. wind up (in set #19) End up in a particular situation, condition or place, esp. an unpleasant one	You could wind up in prison for the rest of your life. 最終的には、残りの人生を刑務所で過ごすことになるかもしれないよ。
82. pull back (in set #19) Move backwards or make SB/STH move backwards	The mother quickly pulled back the child's outstretched hand that was going to touch the electric heater. 母は、電気ヒーターに触れようとしていた子供の手を素早く退けた
102. keep on (in set #19) Continue doing STH without stopping, or repeatedly	I can't keep on spending money if I don't have a job. 仕事がなければ、お金を無駄遣いし続けられない。
20. look up (in set #20) Raise one's eyes	I laid on my back and looked up at the sky. 仰向けになり、空を見上げた。
66. shut down (in set #20) Stop (or make STH stop) working or operating (machine, computer, business, premise, strategy)	In the capital, parts of the subway system were still shut down because of earthquake. 地震の影響で、首都では地下鉄のシステムの一部が未だに停止してした。
83. hang on (in set #20)	"Please hang on a second." Maybelle covered the phone

Wait for a short time	with her hand, but I could hear her yelling. 「少しお待ちください。」メイベルは手で電話を覆ったが、叫んでいるのが聞こえた。
104. make out (in set #20) See or hear with difficulty	She could make out the shape of her mother in the darkness. 彼女は暗闇の中、自分のお母さんの姿を見分けることができた。
44. bring back (in set #21) Make STH/SB return to a place, state, situation, or conversation topic	He was brought back to life in spite of falling under ice water for 10 minutes. 10 分間も氷水に落いていたが、彼は息を取り戻した。
60. go off (in set #21) Go somewhere, esp. for a particular purpose	She'll be going off to graduate school next year to study agribusiness. 彼女は農業関連産業を勉強するため、翌年大学院に行くだろう。
84. build up (in set #21) Increase or cause STH to increase, accumulate, or strengthen, especially progressively	Through the winter, snow built up in layers. 冬の間ずっと、雪が何層にも積み重なった。
100. carry on (in set #21) Continue to do or be involved with STH, or make STH continue (especially despite difficulty)	He studied Latin language and can carry on a conversation in Latin. 彼はラテン語を勉強していたので、ラテン語で会話をなんとか続けることができる。
18. end up (in set #22) Finally do STH or be in a particular place, state, or situation after doing STH or as a consequence of it, esp. unexpectedly	They were arrested and ended up in jail. 彼らは捕まり、最終的には刑務所行きとなった。
58. put down (in set #22) Place STH/SB on the floor or on a flat surface	After seeing that the ink was spilling, I put down the pen and kept reading the article. インクがあふれ出ていたのを見て、そのペンを置いて、その記事を読み続けた。
87. put on (in set #22) Put a piece of clothing or jewelry onto one's body	Since he was nearsighted, he first put on his glasses and then began to read the newspaper. 彼は近眼のため、眼鏡をまず身に着けて、新聞を読み始めた。
99. blow up (in set #22) Explode or destroy STH with a bomb, or cause to be exploded or destroyed	He was arrested for a plot to blow up the US Embassy in Paris. 彼は、パリにあるアメリカ大使館を爆破しようとした陰謀の罪で逮捕された。
8. go out (in set #23) Leave a room, building, car,	She heard the gunshots but couldn't see anything, so she cautiously went out the door into the yard to see what

or one's home to go to a social event	happened. 彼女は銃声を聞いたが、何も見えなかったので、何が起こったのか見るため、ドアから庭へと出た。
43. put up (in set #23) Display or attach STH (e.g., to a wall) so it can be seen	After I put up this big Christmas tree we can all gather around with the Christmas ornaments. この大きなクリスマスツリーを飾った後、私たちはクリスマスの装飾を持ち寄って集まる。
50. move on (in set #23) Start doing or discussing STH new (job, activity, conversation topic)	Science and our understanding of the world have moved on. 科学と世界に対する私たちの理解は、どんどん進んでいる。
85. throw out (in set #23) Refuse to accept or consider (esp. by people of authority)	27, 000 Florida ballots were thrown out because of faulty voting machines. 投票機の不良のために、2 万 7 千ものフロリダの投票は取り消された。
6. find out (in set #24) Discover STH; obtain	People can find out more information on the website. ウェブサイト上では、更に情報を見つけることができる。
45. bring up (in set #24) Raise for discussion or consideration	He brought up the issue of immigration at the conference. 彼は会議で、移民問題を提起した。
49. check out (in set #24) Have a look at; examine STH/SB (esp. to get more information or make a judgement)	Don't forget to check out our Web site, CNN.com, for the latest on this morning's top stories. 今朝のトップ記事の最新情報について、当社の Web サイト「CNN.com」をご覧ください。
76. go through (in set #24) Experience STH difficult or unpleasant	Karen's husband had gone through a period of unemployment. カレンの夫は、失業期間を経験したことがある。

PHaVE Appendix B

This appendix contains example sentences along with the Japanese translations of 60 phrasal verbs which were tested for the delayed post-test of Experiment 1. The sentence and meaning of each phrasal verb were sourced from the PHaVE List (Garnier & Schmitt, 2014), while the Japanese translations were generated by the author. The numerical value accompanying each phrasal verb indicates its rank in the PHaVE List.

Note: An asterisk (*) denotes that the phrasal verbs were included as fixed items in the test.

First semester

Phrasal verb	Sentence
34. look back (in set #1)	We looked back on the past while immersed in memories. 私たちは思い出に浸りながら過去を振り返った。
93. sit up (in set #1)*	The sudden noise made her sit up in her bed and listen. 突然の騒音で、彼女はベッドから起き上がり、耳を傾けた。
103. run out of (in set #1)	We've run out of biscuits. 私たちはビスケットを食べ尽くしてしまった。
50. set off (in set #2)*	We will finish packing and set off in the morning. 私たちは朝のうちに、荷造りを終わらせて出発します。
61. cut off (in set #2)	Take the carrots and cut the ends off (cut out). ニンジンを取って、先端を切り取ってください。
42. look down (in set #3)	She looked down at the ground to see what she stepped on. 彼女は何を踏んだのか確認しようと、地面を見下ろした。
65. clean up (in set #3)*	Make sure you clean up your mess because I won't do it for you. ちゃんと散らかっているものを片付けてね。私はあなたのためにしたくないから。
89. come over (in set #3)	Could you come over and give me a hand with this?

	ちょっと来て、私に手を貸してくれませんか？
72. take back (in set #4)*	After dinner, he took her back to her house. 夕食後、彼は彼女を彼女の家に連れ帰った。
111. stand out (in set #4)	Excellent product quality is what made the brand stand out from its competitors. 優れた製品品質は、そのブランドを競合会社から際立たせている。
19. get back (in set #5)	She got back to London last Monday. 彼女は、先週の月曜日にロンドンへと戻った。
67. turn over (in set #5)	The police officer turned over the criminal to the jail guard. 警官はその犯罪者を刑務官に引き渡した。
96. put back (in set #5)*	Could you put the milk back in the fridge please? 牛乳を冷蔵庫に戻していただけませんか。
46. look out (in set #6)	She liked to go by the window and look out at the garden. 彼女は窓際に行き、窓からその庭を見る（眺める）のが好きだった。
71. line up (in set #6)	Dozens of taxis were lined up at the entrance. 入口には、たくさんのタクシーが並んでいた。
91. start out (in set #6)	She started out (started off) as a shop assistant and gradually climbed the employment ladder. 彼女は店員として仕事を始め、徐々に昇進の階段を昇っていった。
118. hold back (in set #6)	They should not hold back from joining us if they want to. 彼らが望むのなら、私たちに加わることを遠慮する必要はない。
22. sit down (in set #7)	Please sit down and have a drink. 座って飲み物でも飲んでください。
90. move in (in set #7)*	He liked the house so much that he decided to move in immediately. 彼はその家がとてもお気に入りだったので、すぐにそこに引っ越すことにした。
97. send out (in set #7)	Hundreds of copies were sent out (send off) to the local population. 何百ものチラシが地元の人々に送られた。
10. grow up (in set #8)	Seeing my kids growing up is such a lovely thing. 自分の子供が成長していくのを見るのは、とても素敵なことよ。
29. work out (in set #8)	We still need to work out the details of the procedure. 私たちはまだその進行の細かい部分を計画する必要があります。
54. go in (in set #8)*	This restaurant looks really nice; let's go in and have lunch. このレストランはとても良い感じね。中に入って、ランチしましょう。
24. take out (in set #9)*	He tore open the envelope and took out a few bills. 彼は封筒を破って、何枚かの請求書を取り出した。

78. pay off (in set #9)	It will take a dozen years for him to pay off his debts. 彼が借金を返済するのに十数年かかるでしょう。
108. step back (in set #9)	He stepped back when the big man threatened him. その大男が彼を脅したとき、彼は後ずさりした。
32. go ahead (in set #10)	Go ahead and ask me your question! さあどうぞ、質問してください。
86. hang out (in set #10)	I don't like to hang out with people I work with. 私は一緒に仕事をしている人たちと遊びに出かけるのは、好きではない。
12. turn out (in set #11)	Her suspicion turned out to be justified. 彼女の疑いは正当化されたことが判明した。
31. come down (in set #11)*	Come down from the roof or you will hurt yourself. 屋根から降りなさい。さもないと怪我するよ。
13. get out (in set #12)	These prisoners have no hope of ever getting out of jail. 囚人は、刑務所から出ることを望んでいない。
80. break up (in set #12)	Their marriage broke up in 2007. 彼らの結婚生活は、2007 年に終わった。
115. go around (in set #12)*	There is a rumour going around that she is pregnant. 彼女が妊娠しているという噂が広まっている。

Second semester

Phrasal verb	Sentence
14. come in (in set #13)*	I don't have a chimney. So if Santa tries to come in my house, my alarm would sound. うちには煙突がない。サンタが家に入ろうとしたら、警報が鳴るかもしれない。
47. bring in (in set #13)	I brought in my laptop computer today because my office computer is broken. 自分のオフィスのパソコンが壊れているので、今日は自分のパソコンを持ち込んだ。
70. turn up (in set #13)	The search turned up solid evidence against him. その捜査は、彼に対する確かな証拠を見つけ出した。
75. hang up (in set #13)	He hung up the phone without letting her answer his question. 彼は自分の質問を彼女に答えさせずに、電話を切った。
119. write down (in set #13)	You should write down his contact details in case you want to get in touch. 連絡を取りたい時のために、彼の連絡先の詳細を記録しておくべきだ。
5. go back (in set #14)*	He washed the dishes and went back to his room. 彼は食器を洗い、部屋へと戻った。
27. show up (in set #14)	She didn't show up at the meeting. 彼女は、その会合に姿を見せなかった。

64. set out (in set #14)	I set out to discover the truth behind the story. 私はその物語の背後にある真実を見つけ出そうと乗り出した。
92. call out (in set #14)	He could hear a voice call out his name. 彼は自分の名前を大声で呼ばれたのが聞こえた。
116. walk out (in set #14)	She walked out of the meeting feeling irritated by her colleagues. 彼女は同僚に苛立ち、その会合から出て行った。
7. come out (in set #15)	She went into the bank and came out with some money. 彼女は銀行へ行き、いくらか引き下ろした。
28. take off (in set #15)	I took off my shirt and went to bed. 私はシャツを脱いで寝た。
52. look around (in set #15)	They entered the shop and looked around but nobody was there. 彼らは店に入り周囲を見回したが、誰もそこにいなかった。
73. **lay out** (in set #15)	The whole strategy was laid out in detail in a twenty-page document. その戦略の全貌については、20 ページの資料で詳細に説明されていた。
114. break out (in set #15)	Riots broke out that night. その夜、暴動が突然発生した。
9. point out (in set #16)	Experts have pointed out that eating too much sugar is extremely unhealthy. 専門家たちは、砂糖の過剰摂取が健康をかなり害すると指摘している。
30. stand up (in set #16)	He pushed away from the table and stood up. 彼はテーブルを押しのけて、立ち上がった。
56. get off (in set #16)	You need to take the bus and get off at the third stop. そのバスに乗って、ここから 3 番目のバス停で降りてくださいね。
79. hold out (in set #16)	He took the keys and held them out to her. 彼はカギを取って、彼女に差し出した。
105. shut up (in set #16)	Just sit on the chair and shut up! 椅子に座って、黙ってなさい！
16. give up (in set #17)	She had to give up smoking when she got pregnant. 彼女は妊娠した時、喫煙をやめなければならなかった。
26. go down (in set #17)	After hitting the iceberg, the ship began to go down. 氷山に追突した後、その船は沈み始めた。
55. break down (in set #17)	Our car broke down yesterday. 私たちの車は、昨日壊れた。
77. hold on (in set #17)	He held on to his job until the very last day. 彼は最後の最後の日まで自分の仕事を続けた。
106. turn off (in set #17)	People were asked to turn off their phones. 自分の携帯の電源を切るように言われた。

4. come up with (in set #18)	She instantly came up with a solution to the problem. 彼女はすぐにその問題の解決策を思いついた。
33. go up (in set #18)	Oil prices have gone up last year. 昨年、石油の価格が上昇した。
68. slow down (in set #18)	Economic growth has dramatically slowed down. 経済成長は劇的に減速した。
81. bring out (in set #18)	This haircut brings out the natural curl in your hair. このヘアカットは、君の髪の自然なカールを引き出しているね。
98. get in (in set #18)*	The new security lock prevents thieves from getting in. 新しいセキュリティロックは、泥棒の侵入を防ぐ。
3. come back (in set #19)*	She came back to the kitchen with a bottle of fancy wine. 彼女は高級ワインのボトルを持って、台所に戻ってきた。
35. wake up (in set #19)	I was so tired that I woke up at 10 this morning. とても疲れていたので、今朝は 10 時に目が覚めた。
69. wind up (in set #19)	They wound up having to clear off his debts. 彼らは結局、彼の借金を返済しなければならなくなった。
82. pull back (in set #19)	She pulled back the hair from her face. 彼女は、顔の前にあった髪を退けた。
102. keep on (in set #19)	She wiped tears off her cheeks but kept on crying. 彼女は頬の涙をぬぐったが、泣き続けた。
20. look up (in set #20)	He looked up from his book and shook his head. 彼は本から目を上げて、首を横に振った。
39. pull out (in set #20)*	He reached in his pocket and pulled out a gun. 彼はポケットに手を伸ばして銃を取り出した。
66. shut down (in set #20)	You should shut down your computer at night to save electricity. 電気を節約するには、夜間にコンピュータを停止させるべきだね。
83. hang on (in set #20)	Please hang on for a minute, I'll be quick. ちょっと待ってください、急ぎます。
104. make out (in set #20)	I could barely make out his face in the dark. 私は、暗闇の中でも彼の顔をかろうじて識別することができた。
23. get up (in set #21)*	She got up out of her chair and wore her shoes. 彼女は椅子から立ち上がり、靴を履いた。
44. bring back (in set #21)	This will bring back war into the country. これによって、その国に戦争を呼び戻すでしょう。
60. go off (in set #21)	He decided to go off to college. 彼は大学に行くと決めた。
84. build up (in set #21)	Tension was building up among competitors. 競合他社の間で、緊張が高まっている。
100. carry on (in set #21)	I would like to carry on working after I retire. 定年退職後も働き続けたい。

18. end up (in set #22)	She ended up having to sell her car after her accident. 彼女は事故後、自分車を結局売らなければならなくなった。
40. turn around (in set #22)*	She turned around and walked out the door. 彼女は後ろを向いて、ドアを出ていった。
58. put down (in set #22)	She put down her glass and left the bar. 彼女はグラスを置いてバーを去った。
87. put on (in set #22)	You should put on your gloves, it's really cold outside. 外は本当に寒いので、手袋をはめた方がいいよ。
99. blow up (in set #22)	Several attempts were made at blowing up official buildings. 公の建造物を爆破するため、何度も企てられた。
8. go out (in set #23)	We should go out for dinner sometime. いつかディナーに出かけなきゃね。
43. put up (in set #23)	They put up a few posters on the wall. 彼らは、壁に数枚のポスターを貼り付けた。
50. move on (in set #23)	Let's move on to our next topic. 次のトピックに移りましょう。
85. throw out (in set #23)	The president attempted to have the death penalty thrown out. その大統領は死刑を廃止しようとした。
110. bring down (in set #23)*	The rocket attack brought down the airliner. そのロケットによる攻撃は、旅客機を墜落させた。
6. find out (in set #24)	We need to find out who did this to her. 彼女に対して誰がこのようなことをしたのか、知る必要がある。
45. bring up (in set #24)	I didn't think he would bring up the subject. 彼がその議題を提起するなんて、思っていなかった。
49. check out (in set #24)	Check out our website for more information. 詳細については、当社の Web サイトをご覧ください。
76. go through (in set #24)	You have to understand the tough situation she went through before judging her. 彼女について判断する前に、彼女が経験した厳しい状況を理解してあげないと。
120. move back (in set #24)*	We moved back to New York last year. 私たちは昨年、ニューヨークに住居を戻した。

PHaVE Appendix C

This appendix contains sentences and their corresponding Japanese translations for 120 phrasal verbs that were utilized in *Moodle* activities for Experiment 2. Each phrasal verb is typically associated with a single definition representing the most commonly used meaning according to the PHaVE List (Garnier & Schmitt, 2014). However, in cases where there were insufficient examples for the most common definitions or when the first and second definitions significantly overlapped, both definitions were included. The number assigned to each phrasal verb corresponds to its rank in the PHaVE List.

Note: STH refers to "something," and SB refers to "somebody."

First semester

Phrasal verb and its meaning	Sentences
1. go on (in set #1) Happen, take place	There has been a lot of terrorism going on throughout the world. 世界中でたくさんのテロが起こっている。
	I have no idea what is going on in New York. ニューヨークで今何が起こっているのか私には全く分からなかった。
34. look back (in set #1) Think of STH again; reconsider STH past	Looking back two years ago, I had never thought of becoming a doctor. 2年前を振り返ると、自分が医者になるなんて思ってもみなかった。
	Let's look back at the history of American politics at this time. この時代のアメリカの政治の歴史について振り返ってみましょう。
45. sit up (in set #1)	He suddenly sat up in bed with tears in his eyes.

Rise from a lying to a sitting position	彼は涙を浮かべながら突然ベッドから起き上がった。
	Many people sat up all night watching the Olympics. たくさんの人がオリンピックを見ようと、夜遅くまで起きていた。
63. pull up (in set #1) Stop or cause a vehicle to stop	A police car pulled up and two cops leaped out. 警察車が止まり、2 人の警官が飛び出してきた。
	I asked a cab to pull up to the front of the hotel at 6 a.m. 午前 6 時にホテル前にタクシーを止めるようお願いした。
103. run out of (in set #1) Use STH (or become used) completely so that nothing is left	We are going to run out of food, so I need to go shopping within a few days. 食料が尽きてしまいそうなので、数日のうちには買い物に行かないと。
	I'm afraid we're going to run out of money for fuel before winter's over. 冬が終わる前に燃料費が底を突いてしまうことを懸念している。
2. pick up (in set #2) Get or take SB/STH from a place	I picked up the phone and called my friend. 電話を取って、友人に電話をした。
	I stopped by the store to pick up the reserved book. お店に寄って予約した本を取ってきた。
37. take over (in set #2) Gain control, management, or possession of STH/SB	She took over the company from her father at 27 years old. 彼女は 27 歳の時、父から会社を引き継いだ。
	We shouldn't be worried about AI taking over the world. AI が世界を支配するなどと悲観するべきではない。
50. set off (in set #2) Start on a trip or journey	Christina hired a car and set off immediately. クリスチーナは車を借りて、すぐに出発した。
	Bill had fun at the dance, and set off for home around 10 p.m. ビルはダンスを楽しみ、午後 10 時ごろに家へと出発した。
61. cut off (in set #2) Remove a part of STH by cutting it	She cut off his hair while he was sleeping. 彼女は彼が寝ている間に彼の髪を切った。
	One of his arms was cut off in the accident a decade ago. 10 年前、その事故で彼の片腕が切断された。
94. turn down (in set #2) Refuse or dismiss	The singer has never turned down a request for an autograph or a picture from her fans. その歌手は、ファンからのサインや写真のお願いを決して断ったことがない。
	He turned down an offer to attend at the inauguration ceremony. 彼はその就任式への出席の申し出を断った。
11. set up (in set #3) Establish or create STH; arrange for STH to happen or exist	They will set up an ice cream shop in Central Park this summer. 彼らはこの夏セントラルパークでアイスクリームのお店を設置することにした。

	Would you like to set up a meeting time with him? 彼との面談時間を設定しますか？
42. look down (in set #3) Lower one's eyes to see what is below	He didn't reply, looking down at his hands. 彼は自分の手を見下ろして、返事をしなかった。
	I stood at the window and looked down the ground. 私は窓辺に立ち、グラウンドを見下ろした。
65. clean up (in set #3) Get rid of dirt, mess, pollution, or chemical substances in a place or area	I often clean up the kitchen after we eat when I return to home. 帰省した時には時折、食事後にキッチンを掃除する。
	Her parents hadn't cleaned up her room because they knew she would be back in a few months. 両親は彼女が数か月後には戻ってくることを知っていたので、部屋を掃除しなかった。
89. come over (in set #3) Come to a place or area, esp. towards SB or to join SB	She came over to the house and played with my kids. 彼女が家にやって来て、子供たちと遊んでくれた。
	I think he will come over for dinner, if you invite him. 君が彼を招待したら、彼は食事会にやって来ると思うよ。
113. play out (in set #3) Happen or develop; be enacted or performed	A life-and-death drama was played out on small stage. 生死のドラマが小さなステージ上で演じられた。
	I do never want to see the dreadful scene that was played out in the horror movie. そのホラー映画で展開された恐ろしいシーンは二度と見たくない。
15. take on (in set #4) Undertake or handle	He took on a role as a global ambassador for the American gas and oil industry. 彼はアメリカの石油ガス産業担当大使という役目を引き受けた。
	He has taken on the task of serving a wide range of Asian dishes. 彼は様々なアジア料理を提供する仕事を引き受けている。
38. hold up (in set #4) Hold STH in a high position	Hank cleared his throat and held up his hand to draw attention. ハンクは咳払いをして、注目を集めるために手を挙げた。
	"How long does the repair takes?" Holding up three fingers, the repairer said, "Almost three weeks." 「修理にどのくらいかかりますか？」その修理工は指を三本挙げて、「約 3 週間です。」と答えた。
72. take back (in set #4) Take STH/SB to a place, or time period, they were in before	Donald Trump praised the Brexit vote, saying it indicates the British people "taking back their country." ドナルド・トランプ大統領は、英国の人々は「自分たちの国を取り戻している」ことを示していると言って、Brexit 投票結果を称賛した。
	Iraqi forces fight to take back the city of Mosul from ISIS. イラク軍は ISIS からモスルの都市を取り戻すために戦っ

	ている。
95. back up (in set #4) Move or drive backwards a short way	She backed up the car inside the garage. 彼女は車庫の中で車を後ろに移動させた。
	Heavy rain caused rain water to back up three to six feet on roads and intersections. 豪雨の影響で、道路や交差点で３～６フィート雨水を逆流させた。
111. stand out (in set #4) Distinguish oneself/itself by being better, more significant or more impressive than other people/things	The girl was tall, with long dark hair standing out in the crowd. その少女は身長が高く、人ごみの中でも長い黒髪が際立っていた。
	I saw the thick veins standing out on his arms. 彼の腕の太い血管が浮き出てきたのが見えた。
19. get back (in set #5) Return to a place, position, state, activity, conversation topic	Get back to work, Charlie. Here's a list of unfinished business. 仕事に戻って、チャーリー。これがまだ終わっていない仕事のリストよ。
	It's too cold to wait outside. Let's get back in the car to stay warm. 外で待つには寒すぎるよ。車に戻って温まろう。
48. open up (in set #5) Make STH become available or possible, less limited	Meteorite impact killed off the dinosaurs and opened up developmental opportunities for the distant ancestors of human beings. 隕石衝突は、恐竜を全滅させ人間の遠い先祖が発展する機会を切り開いた。
	The digital technology has opened up more possibilities with visualizations. デジタル技術は、視覚化によってさらなる可能性を広げている。
67. turn over (in set #5) Surrender possession or control to SB/STH	The information was turned over to special prosecutor. その情報は特別検察官に引き渡された。
	The military turned over power to elected civilian government. *civilian government: 民政 軍は、選挙で選ばれた民政に権力を引き渡した。
96. put back (in set #5) Move STH/SB to a place, position, or state they were in before	After the children finished playing, their mother put the room back in order. 子供たちが遊び終わった後、母親は部屋を元の状態に戻した。
	We hope the agreement will be put back on track. 私たちはこの合意が再び軌道に戻ることを願っている。
117. get through (in set #5) Succeed in reaching a physical destination or	In his wheelchair it's hard to get through the door without help. 車椅子では、助けなしでそのドアを通り抜けることは難しい。
	It took her a long time to get through the big crowd to the gate.

stage	その大混雑を通り抜けて出入口に辿り着くのにとても時間がかかった。
21. figure out (in set #6) Come to understand or determine STH	The teacher-librarian serves as a guide to help students figure out the answer on their own. その司書教諭は、生徒たち自身で答えを見つけ出せるように導く仕事をしている。
	By studying the event horizon through both theory and observation, scientists could soon figure out how the universe began. *event horizon: 事象の地平面 研究者は、事象の地平面を理論や観測を通して研究することで、どのように宇宙が始まったのかを発見することができた。
46. look out (in set #6) Look outside, or at the horizon	She was looking out the window and enjoying the autumn leaves. 彼女は窓から外を見ながら、紅葉を楽しんでいた。
	In the morning, I heard this banging on the door. And I looked out the door and all I saw was cops. 今朝、激しくドアを打つ音が聞こえたので、ドアから外を見ると、全員警察だった。
71. line up (in set #6) Form or make SB/STH form into a line	Hundreds of fans lined up in London to see the first showing of “Harry Potter and the Cursed Child.” 何百ものファンたちは、ロンドンで「ハリーポッターと呪いの子」の封切りを観るために並んでいた。
	A few cars had lined up on the road, waiting to get through the intersection. その交差点を抜けるために待つ数台の車が、道路上で並んでいた。
91. start out (in set #6) Start a life, existence, profession, or course of action in a particular way or by doing a particular thing	My day had started out to be a fantastic one, but it had gone downhill quickly. 素晴らしい私の 1 日が始まったけど、あっという間に下り坂だった。
	Sulivan started out his career as a lawyer fighting cybercrime. サリバンは、ネット犯罪と闘う弁護士としての生涯を始めた。
118. hold back (in set #6) Decide not to do or say STH	The politicians made efforts to hold back their negative comments until the final vote. 政治家たちは、決選投票が終わるまで否定的なコメントを控えるように努力した。
	I couldn’t hold back information from these two people. They were terrified. その 2 人からの情報を黙っておけなかった。彼らに脅されたんだ。
22. sit down (in set #7) Move from a standing position to a sitting	She removed a stack of mail from a chair and sat down. 彼女は椅子から大量の手紙をどけて、座った。
	When he played the piano or sat down for dinner, he always

position	wore a tie. 彼はピアノを弾く時や夕食の卓に着く時でさえも、ネクタイを付けていた。
36. carry out (in set #7) Perform or complete	The bomb attacks were carried out by terrorists. その爆弾による攻撃は、テロリストによって実行された。
	The medical research was carried out by famous doctors. その医療研究は、有名な医師たちによって行われた。
53. be/get caught up with (in set #7) Become involved in STH which prevents SB from making progress or moving forward	Her long hair was caught up in a clip at the nape of her neck. 彼女の長い髪は、襟元のクリップに巻き込まれていた。
	Many children are getting caught up in their parents' quarrels. 多くの子供たちは親の喧嘩に巻き込まれている。
90. move in (in set #7) Settle into a new house or place	I just left the house and moved in with my parents. 私はその家を出て、両親と一緒に新しい家へと引っ越した。
	That Italian family moved in across the street decades ago. そのイタリア人の家族は、何十年か前に通りの向こう側に引っ越してきた。
97. send out (in set #7) Mail, send or distribute to a number of people	It is illegal to send out jamming signals. *jamming: 電波妨害 妨害信号を発信することは、法で禁じられている。
	The first six months, I sent out 200 letters requesting interviews, and they were all rejected. 最初の 6 か月間、インタビューをお願いするために 200 通もの手紙を送ったが、全て断られた。
10. grow up (in set #8) Gradually advance in age and maturity	Sadly, there are still many Indian children who are growing up in poor environments. 悲しいことに、今も貧しい環境の中で育つ多くのインドの子供たちがいる。
	Lots of kids in this town grow up, wishing one day they could play for Big Red football. この町の多くの子供たちは、いつか Big Red Football でプレイできたらと願いながら、大人になっていく。
29. work out (in set #8) Plan, devise or think about STH carefully or in detail	The details will be worked out in subsequent meetings. 詳細は次のミーティングで計画されるだろう。
	A deal has been worked out to help resolve Europe's debt crisis. 欧州金融危機を解決するために、政策が考え出された。
54. go in (in set #8) Enter	He opened the door and went in to the room. 彼はドアを開けて部屋に入った。
	Why don't we go in/into the house so that we can put these kids to bed? 子供たちを寝かせるために、家の中に入りましょう。

88. get down to (in set #8) Begin to pay serious attention to STH	New York City's new mayor got down to business, but there was a battle over his plan. ニューヨークの新市長は事業に取り組んだが、彼の計画をめぐって闘争があった。
	The newly appointed man in charge is wasting no time to get down to work. 新しく指名された担当は、仕事に取り組む時間を無駄にしない。
109. lay down (in set #8) Put STH away or down on a surface, esp. because one has stopped using it	The criminal laid down his gun and raised his hands when the police surrounded him. 警官に取り囲まれると、その犯人は銃を置いて両手を挙げた。
	The separatists in the country will lay down their arms. その国の分離主義者たちは、武器を捨てて戦いをやめるだろう。
24. take out (in set #9) Remove STH/SB from somewhere	Dylan reached into his pocket and took out a business card. ディランはポケットに手を突っ込んで、名刺を取り出した。
	I have to sweep floors, vacuum carpets, and take out trash. 床を掃いて、カーペットに掃除機をかけて、ゴミを出さなきゃ。
25. come on (in set #9) Said to encourage SB to try harder, or do or say STH	Don't be silly, Sylvia. It was a long time ago. Come on. You don't have to talk to him. ふざけないでよ、シルビア！ずっと前のことじゃない。しっかりして。彼に話す必要なんてないよ。
	Come on, mom. Don't fall asleep. しっかりして、ママ。寝ないでよ。
51. put out (in set #9) Make STH known or accessible to the public	A big responsibility falls on the media to put out accurate information. 正確な情報を公開する重大責任がメディアにはある。
	President Putin put out a statement today that this fake news was indeed fake news. プーチン大統領は今日、このフェイクニュースは本当に偽物の報道であるという、声明を公開した。
78. pay off (in set #9) Pay the complete amount of STH	Today's student loan borrowers have struggled to pay off debt over the past decade. 今日の奨学金を借りた人たちは、過去 10 年以上もかけてその借金を完済するのに苦労している。
	You can pay off your mortgage when you retire. 君が定年退職した時に、住宅ローンを完済できるよ。
108. step back (in set #9) Move back by lifting one's foot and putting it down backwards	I hesitated to go outside for a moment, then stepped back from the doorway. 外に出るのを少しの間ためらって、玄関口から後ろに下がった。
	She knocked on the door and then stepped back so he wouldn't

	be afraid. 彼女はドアをノックして、彼が怖がらないように一歩後ろに下がった。
17. make up (in set #10) Form the whole of an amount or entity	African American students make up 80 percent of the student population in this city. アフリカ系アメリカ人の学生は、この市の学生の 80 パーセントを占めている。
	The group then was made up of classmates from Howard University's jazz program. そのグループは、ハワード大学のジャズの講義を受けているクラスメイトで構成されていた。
32. go ahead (in set #10) Proceed with a course of action without further hesitation	A: Let me ask you a question. B: Go ahead. A: 質問させてください。 B: そのまま続けて。
	A: This is Steve. B: Hi, Steve. You're on the air now. Go ahead, please. A: こちらはスティーヴです。B: こんにちはスティーヴ。現在生放送中です。どうぞ話を進めて。
62. turn back (in set #10) 1) Turn around so as to face the opposite direction 2) Go back (or make SB/STH go back) in the direction SB/STH has come from	He shut the door quickly and turned back to his friend. 彼はドアを素早く閉めて、後ろにいた友人の方を向いた。
	I have a lot of regrets. If I could, I would turn back time. 自分にはたくさんの公開がある。できるなら、時間を戻したい。
86. hang out (in set #10) Spend time relaxing or enjoying oneself	I will go to my guitar lesson and hang out with my friends today. 今日はギターの授業受けて、友達と遊ぶ予定です。
	The kids hung out in the art room. その子供たちは、美術室で楽しい時間を過ごした。
112. come along (in set #10) Appear or arrive; come into existence	Humans came along a hundred thousand years later than scientists had originally thought. 人類は、科学者たちが最初に考えていたよりも 10 万年も遅く誕生していた。
	If some other opportunity comes along, we'll be ready to seize it. もし他の機会が訪れたら、それをとびつく準備はできている。
12. turn out (in set #11) Prove or be discovered to happen or be	Sometimes your life does not turn out the way you expected. 人生は、自分が期待したようになるわけではない。
	The story turned out differently from what we had expected. その物語は私たちが予想していたものとは違うということが分かった。
31. come down (in set #11)	When he appeared on stage, the rain came down even harder. 彼がステージ上に現れた時、さらに強く雨が降ってきた。

Move from a higher spatial location to a lower one; fall/land onto the ground	She came down the stairs and crept into the room. 彼女は階段を下りて、部屋にそっと入っていった。
57. keep up (in set #11) Move, progress or increase at the same rate or pace as SB/STH	We have to create 150,000 jobs every month just to keep up with population growth. 人口増加に遅れを取らないために、毎月 15 万人分の雇用を創出しなければならない。
	Amazon - the world's largest bookseller - is now unable to keep up with demand. 世界最大の書籍を販売している Amazon は今や、需要に追いつけていない。
74. go over (in set #11) Move towards a place or person, esp. by crossing an area	You shouldn't go over the line with Internet friends. ネット上の友達とは一線を超えるべきではない。
	I captured a photo of the locomotive going over the bridge. 私はその機関車が橋を越えていく写真を撮った。
107. bring about (in set #11) Cause to happen or emerge, esp. STH positive	War itself can never bring about true peace. 戦争は決して本当の平和をもたらさない。
	What brought about the change in his attitude toward me? 何が私に対する彼の態度を変化させたの？
13. get out (in set #12) Leave a container or make SB/STH leave a container	Japanese people make the bad spirits get out of the house by throwing dried soybeans on Setsubun. 節分の日には、日本人は豆をまいて厄を家から追い出す。
	She left the car running and got out. 彼女は車を作動させたまま、出てきた。
41. take up (in set #12) Use a particular amount of space, time or effort	Much of the time was taken up with a discussion. 多くの時間が議論に費やされた。
	The rest of the space was taken up by a grand piano. 残りのスペースはグランドピアノが陣取っていた。
59. reach out (in set #12) Stretch an arm in order to hold, touch, or get STH that is within short distance	The baby's tiny hand reached out to touch his mother's face. その赤ちゃんは、お母さんの顔に触れようと自分の小さな手を伸ばした。
	He reached out his arms to hold her. 彼は彼女を抱きしめるために、腕を伸ばした。
80. break up (in set #12) End or cause to end or fail	The men wrestled until the fight was broken up by the police. その男たちは、警官が闘いを終わらせるまで取っ組み合いをしていた。
	The meeting broke up in disagreement, with other issues on the backburner. そのミーティングは、他の問題で後回しにされ、不一致で終わった。
115. go around (in set #12) Go from one place/person to another; circulate	I had an opportunity to go around the country and speak to a lot of various people. 私は国中を周り、多くの様々な人と話をする機会があった。
	I know some of you, but let's go around the table. Give us

	your name and tell us what you think about this project. 君たちの中には知っている人もいますが、一人ずつ伺いましょう。自分の名前とこの計画についてどう考えているのかを教えて下さい。

Second semester

Phrasal verb and its meaning	Sentences
14. come in (in set #13) Enter a place or area (room, building)	We had pizza delivered, but the delivery person didn't know which door to come in. ピザの配達を頼んだけど、配達員はどのドアに入れば良いか分からなかった。
	Why are you here? You mustn't come in the room! It's not allowed! 何故ここにいるの？部屋に入らないで！許可されてないよ！
47. bring in (in set #13) Bring STH to a place or situation	According to public documents, the company brings in around a hundred million dollars a year. 公的文書によると、その会社は 1 年におよそ 1 億円の利益をもたらしている。
	WhatsApp currently does not sell advertising and brings in little revenue. WhatsApp というアプリは現在、広告の営業をしておらず、ほとんど収入をもたらしていない。
70. turn up (in set #13) Yield; be (or make STH be) found, discovered, or noticed	A quick search turned up a recipe and picture from some unnamed cookbook. 検索すると即時に、ある不特定の本からレシピと写真が見つかった。
	The examination of Mr. Martin's computers has so far turned up no evidence that it was hacked. マーティンさんのコンピュータの調査では、今のところハッキングされた証拠は見つかっていない。
75. hang up (in set #13) Finish a conversation on the telephone by putting the receiver down or switching the phone off	Sienna blew a kiss through the phone and hung up. シエナは電話に投げキスをして、電話を切った。
	The caller hung up without responding. 電話をかけてきた人は、何も返事をせず電話を切った。
119. write down (in set #13) Record information on paper	The detective wrote down the names of all his brothers and sister. その探偵は、彼のすべての兄弟姉妹の名前を書き留めた。
	The officer wrote down the number plate, inspecting the windscreen damage.

	その警官は、車のフロントガラスの破損を検査し、その車のナンバープレートを記録した。
5. go back (in set #14) Return to a place, time, situation, activity, conversation topic	Kids started to go back to school after a long vacation. 子供たちは長期休みが終わり、学校に戻り始めた。
	I went back to the house earlier to watch a soccer game on TV. テレビでサッカーの試合を見るために、今日は早く家に帰った。
27. show up (in set #14) Make an appearance at a social or professional gathering	Roughly 1,300 fans showed up to watch the opening game of baseball. およそ 1300 人のファンが野球の開幕戦を見るために、姿を見せた。
	I called 911, and two hours later the cops showed up, made a few notes, and went away. 911 に電話して 2 時間後、警官が姿を現し、メモを少し取って、立ち去った。
64 go back (in set #14) Start doing or working on STH, esp. with a particular goal in mind	The study set out to investigate the factors that impede primary school attendance in Nigeria. その研究は、ナイジェリアの小学校への就学を妨げている要因の調査へと乗り出した。
	The Iranian government set out to block the satellite network system. イラン政府は、衛星通信ネットワークシステムの妨害に乗り出した。
92. call out (in set #14) Speak or utter loudly	I found myself shutting my car door and calling out his name. 気付いたら車のドアを閉めて、彼の名前を叫んでいた。
	Sometimes while I am dreaming he calls out the words “I don’t know” over and over again. 時折、彼が何度も何度も「知らない」という言葉を叫んでいる夢を見る。
116. walk out (in set #14) Leave a place or event, especially suddenly or angrily	Davis walked out of prison, a man with a second chance. デイビスはやり直すチャンスをもらい、刑務所から突然と出て行った。
	The gunman suddenly walked out the back of the building and killed four people. その銃器を持った男は突然、建物の裏から出てきて、4 人をも殺害した。
7. come out (in set #15) Leave a place (room, building, container) or appear from it	He opened his mouth but no words came out. 彼は口を開けたが、何も言葉が出てこなかった。
	People are always surprised to hear the big voice coming out of the little kid. いつも、その小さな子供から出てくる大きな声を聞いて驚かれる。
28. take off (in set #15)	He took off his clothes and got in the shower.

Remove STH (esp. piece of clothing or jewelry from one's body)	彼は服を抜いで、シャワーを浴びた。
	He took off his hat, took a handkerchief from his jacket pocket, and wiped the sweat from his forehead. 彼は帽子を脱いで、ハンカチをジャケットのポケットから取り、額の汗をぬぐった。
52. look around (in set #15) Examine a place or one's surroundings so as to view what it might contain or look for a particular thing	She looked around the table and asks the students to introduce themselves one by one. 彼女はテーブルを見回し、生徒に一人一人自己紹介をしてもらった。
	The tiger opened his eyes and looked around until he found his mother. そのトラは目を見開いて、お母さんを見つけるまで辺りを見回した。
73. lay out (in set #15) Describe or explain STH clearly or in detail, esp. officially and in writing	He is laying out a vision to make America great again. 彼は、再びアメリカを素晴らしいものにしようとするビジョンを明確に提示している。
	The President has laid out the priorities of what should be in healthcare reform. その大統領は、医療改革において何をすべきかの優先事項を明確に説明した。
114. break out (in set #15) Start suddenly, esp. STH undesirable and unpleasant	When World War I broke out in 1914, Gilson, Smith and Wiseman joined the Army. 1914 年に第一次世界大戦が勃発した時、ギルソン、スミス、そしてワイズマンは陸軍に加わった。
	He told everyone to leave when a fight broke out. Another fight started outside, and he heard three shots fired. 彼は喧嘩が起こった時、皆に出て行くように言った。また別の喧嘩が再び外で始まり、彼は 3 発の銃弾が発砲されるのを聞いた。
9. point out (in set #16) Direct attention toward STH (fact, idea, information)	Critics immediately pointed out the outrageous nature of the show. 批評家はすぐに、そのショーの異常性について指摘した。
	City officials have pointed out that kids now often play the same sport all year around. 市の職員は、今子供たちは一年中同じスポーツをすることが多いと指摘している。
30. stand up (in set #16) Rise to a standing position after sitting or lying down	In this concert, we have to stand up, give a yell, and jump! このコンサートでは、立ち上がって、叫んで、ジャンプしなきゃ！
	His sandy hair is standing up in all directions. 彼の薄茶色の髪は、四方八方に立っている。

56. get off (in set #16) Go away from, leave (train, bus, aircraft, lift)	When I get off the bus the weather has changed again and it's barely raining anymore. バスを降りると、天気は再び変わり、もう雨は降っていない。
	I wouldn't know when to get off the train without the driver making an announcement. 運転手がアナウンスをしてくれなければ、電車をいつ降りればよいか分からなかっただろう。
79. hold out (in set #16) Move one's hand or an object in one's hand forward or towards SB, in order to grab or give STH	Valerie held out her arms and Laila hugged her. ヴァレリーは腕を伸ばし、ライラは彼女にハグをした。
	"Do you know what these things are called?" the boy asked, holding out his palm with the ant on it. 「これなんて呼ぶか知ってる？」と、アリを乗せた手のひらを差し出して、少年が聞いてきた。
105. shut up (in set #16) Stop (or make SB/STH stop) talking or making a noise	Will you guys shut up? We're trying to concentrate on studying. 黙ってくれないかな？今勉強に集中しようとしているんだ。
	The dog immediately started barking again, but he shouted through the door at it, "Shut up, dog." その犬は、またすぐに吠え始めたが、彼はドア越しに「黙ってくれ」とその犬に向かって叫んだ。
16. give up (in set #17) Stop doing or having STH; abandon (activity, belief, possession)	He hasn't given up hope for a remedy that might extend his life. 彼は延命できるかもれない治療への希望を捨てなかった。
	I was injured physically and it took eighteen months to make a full recovery. I had to drop out of police college and give up the job waiting for me. 私は身体的に傷を負い、完治するまでに 18 か月もかかった。警察学校を退学し、私を就くはずの仕事を諦めなければならなかった。
26. go down (in set #17) Move down to a lower level or position	The sun was going down behind the mountains. 太陽は山々の後ろへと沈んでいった。
	She went down the stairs of the subway. 彼女は地下鉄の階段を下りて行った。
55. break down (in set #17) Stop working or functioning; fail or collapse (vehicle, device, relationship, negotiations)	When the negotiations broke down last week, the tone turned hostile. 先週交渉が上手くいかなくなった時、口調が敵意あるものになった。
	The immune system has to distinguish between self and nonself. Sometimes this intricate system breaks down. 自己免疫か非自己免疫かを区別するのが、免疫システムである。時折、この複雑なシステムは上手く働かなくなる。
77. hold on (in set #17) Refuse to let go of STH	In the Middle East no leader can hold on to power for long without oil.

	中東では、先導者は石油なしでは権力を持ち続けられない。
	I tried to catch the woman who was about to slip onto the train tracks, but she couldn't hold on to my hand and she fell. 電車の線路上に滑り落ちそうな女性をつかもうとしたが、私の手をつかめず、そのまま落ちてしまった。
106. turn off (in set #17) Stop a piece of equipment working temporarily or a supply flowing by turning a tap, pressing a button, or moving a switch	I turn off the TV and listen to music before sleeping. 寝る前には、テレビを消して音楽を聴く。
	I always become worried if I turned off lights after leaving my house. いつも、家を出た後に電気を消したか心配になる。
4. come up with (in set #18) Bring forth or produce	When he came up with ideas, he always had a plan. 彼が考えを思いつく時には、常に計画があった。
	The group probes child homicides and unexplained child deaths, and strives to come up with ways to prevent such tragedies. その団体は、子供の殺人事件や原因不明の子供の死を調査し、このような悲劇を防ぐための方法を考え出そうと努力している。
33. go up (in set #18) Become higher in value; increase	Either taxes will go up or the quality of services, such as health care, will be degraded. 税金が上がるか、保険医療のようなサービスの質が低下するかのどちらかである。
	Premiums are supposed to go up 15 percent to 20 percent. 保険料は 15 パーセントから 20 パーセントに上がるはずである。
68. slow down (in set #18) Move, proceed or progress at a slower pace (vehicle, economy)	As acidity increases, shells become thinner, growth slows down and death rates rise. 酸性度が上がると、貝は厚みがなくなり、成長は遅れ、死亡率が増加する。
	After losing 3 to 5 percent of our body weight, the sweating process began to slow down. 3 から 5 パーセント体重を落とした後、汗をかく過程は遅くなり始めた。
81. bring out (in set #18) Make a particular detail, quality or feeling more noticeable than it usually is	I put a tiny bit of red pepper. It's not spicy but it brings out the flavor. ほんの少量のトウガラシを入れます。辛くなく、香りを引き立ててくれます。
	I finished the paintings, using lines of varying thickness to bring out details. 細かさを引き立てるために、線の太さを変化させながら、絵を描き終えた。
98. get in (in set #18) Go (or make STH/SB go)	If I wanted a good seat I should get in line. 良い席を取りたければ、列に入らないと/並ばないとい

inside a place (car, house, room)	けないよ。
	She walked to my car, got in, and drove away. 彼女は車の方へ歩いて、中に入り/車に乗り、去っていった。
3. come back (in set #19) Return to a place or conversation topic	He broke a number of bones and didn't come back to school. 彼は何ヵ所も骨折して、学校に戻ってこなかった。
	We're going to take a little break and come back and talk about some of the other issues. 少し休憩して、皆が戻ってきてから、ほかの問題についていくつか話しましょう。
35. wake up (in set #19) Become (or make SB become) conscious again after being asleep	This morning, I woke up from a strange dream. 変な夢のせいで目が覚めた。
	She slept for a few hours at night and then woke up and started to chat with her friend on the Internet. 彼女は夜中に数時間寝て、目を覚まし、インターネット上で友達と雑談し始めた。
69. wind up (in set #19) End up in a particular situation, condition or place, esp. an unpleasant one	She got arrested for drug possession and wound up in jail. 彼女は麻薬所持で逮捕され、最終的に刑務所入りとなった。
	It's easy to wind up wasting money on entertainment. 結局、娯楽は簡単にお金の無駄使いで終わってしまう。
82. pull back (in set #19) Move backwards or make SB/STH move backwards	She pulled back the curtain to see the sunrise. 日の出を見るために、カーテンを退けた。
	Her curly black hair is pulled back in a ponytail. 彼女の黒い巻き毛は、後ろの方でポニーテールにまとめられている。
102. keep on (in set #19) Continue doing STH without stopping, or repeatedly	I can't stand it anymore. I can't keep on enjoying life. もうこれ以上耐えられない。人生楽しみ続けられないよ。
	Can the business keep on expanding without hiring more people? 更に雇用せずに、事業は止むことなく広がり続けられるのだろうか？
20. look up (in set #20) Raise one's eyes	He blinked, opened his eyes, and looked up at the moon. 彼はまばたきして、目を見開いて、月を見上げた。
	Samantha tilted her head back to look up at her mother. サマンサは頭を後ろに傾けて、母を見上げた。
39. pull out (in set #20) Take STH/SB out of a container, thing or place	He pulled out his wallet and gave twenty dollars to pay the bill. 彼は財布を取り出し、会計で 20 ドルを渡した。
	She opened the bag and pulled out a pen and piece of paper. 彼女はカバンを開けて、ペンと紙を取り出した。
66. shut down (in set #20) Stop (or make STH stop) working or operating	A lot of companies are shutting down non-essential operations for saving. 多くの会社は、節約のために不必要な機械の運転を停止

(machine, computer, business, premise, strategy)	している。
	The storm has shut down electric power for the entirety of Puerto Rico. その嵐はプエルトリコ全体の電力を停止させた。
83. hang on (in set #20) Wait for a short time	"Do you have any paper towels in the kitchen?" "Oh. Yeah, sure. Hang on a minute." 「キッチンにキッチンペーパーある？」「うん、あるよ。ちょっと待ってね。」
	Hang on a moment. Don't leave, have another cup of coffee. ちょっと待ってて。どこか行かずにもう 1 杯コーヒーでも飲んでてね。
104. make out (in set #20) See or hear with difficulty	He tried to speak, but I couldn't make out his words. 彼は喋ろうとしたが、私は彼の言葉を理解することができなかった。
	The light was insufficient to make out the features in the painted faces in the portraits that lined the walls. 壁に並んでいる肖像画の描かれた顔の特徴を見分けるのに、明かりが足りない。
23. get up (in set #21) Rise or cause to rise after lying in bed or sitting/kneeling	When I could not sleep, I got up from my bed and watched TV. 眠れないときには、ベッドから起きてテレビを見た。
	The older we get, the more we get up at night to use the restroom. 歳をとると、夜中に起きてトイレに行く回数が増える。
44. bring back (in set #21) Make STH/SB return to a place, state, situation, or conversation topic	The misstep brought back memories of my own freshman year. その失敗は、新人時代の記憶をよみがえらせた。
	I was there during the war. So the movie brings back a lot of images of the war. 戦時中そこにいたので、その映画はたくさんの戦争の映像をよみがえらせる。
60. go off (in set #21) Go somewhere, esp. for a particular purpose	At seventeen, Jinx went off to college on scholarship. 17 歳のころ、ジンクスは奨学金で大学へ行った。
	He is unsure whether to go off to school or join the family boat business. 彼は、学校へ行くか家族の船の事業に加わるか、まだ分からない。
84. build up (in set #21) Increase or cause STH to increase, accumulate, or strengthen, especially progressively	We finally released the pressure that had built up for so many years. 私たちは遂に、何年も蓄積していたプレッシャーを解き放った。
	Roosevelt took advantage of Long's inactivity to build up naval forces.

	ルーズベルトは海軍を増強しようと、ロングの無活動状態をうまく利用した。
100. carry on (in set #21) Continue to do or be involved with STH, or make STH continue (especially despite difficulty)	Maria and her sister Luisa carried on the work started by their parents. マリアと彼女の妹ルシアは、両親が始めた仕事を続けた。
	Clive tried to carry on business as usual, but everyone knew he was still hurting. クライヴはいつも通り事業を続けようとしていたが、みんな彼がまだ傷ついていることを知っていた。
18. end up (in set #22) Finally do STH or be in a particular place, state, or situation after doing STH or as a consequence of it, esp. unexpectedly	She ended up in prison for robbing a credit card. 彼女はクレジットカードを盗んだ罪で、最終的に刑務所入りとなった。
	When you play in casino, you end up spending a large amount of money. カジノで遊ぶと、結局大金を使ってしまうことになるよ。
40. turn around (in set #22) Go from one place/person to another; circulate	Chuck felt a tap on his shoulder and turned around. チャックは肩を軽くたたかれたのを感じ、後ろを向いた。
	At the door, she turned around and looked at him again. ドアの前で、彼女は後ろを向いて彼をもう 1 度見た。
58. put down (in set #22) Place STH/SB on the floor or on a flat surface	He put down the paper and began to speak. 彼は、紙を置いて話し始めた。
	When Gersh put down the phone, her hands were shaking. ガーシュは電話を置いた時、手が震えていた。
87. put on (in set #22) Put a piece of clothing or jewelry onto one's body	Sasha put on clothes and went outside without a word. サーシャは衣服を身につけて、一言も言わず出て行った。
	I heard an admonition sounded – "Put on your shoes!" from inside the house - and there was the boy and his mother. 「靴を履きなさい！」と家の中から説教している声が聞こえてきた。そこには、少年とお母さんがいた。
99. blow up (in set #22) Explode or destroy STH with a bomb, or cause to be exploded or destroyed	His stories were creative and consisted of dragons, blowing up the world. 彼の物語は独創的で、世界を破壊するドラゴンで構成されている。
	A car bomb was blown up outside a busy restaurant. 込み合っているレストランの外で、車の爆弾が爆破された。
8. go out (in set #23) Leave a room, building, car, or one's home to go to a social event	I will go out for dinner with his family today. 今日は彼の家族と夕食に出かける。
	I want to just go out this night and take pictures of night viewing. 今夜出かけて、夜景の写真を撮りたい。
43. put up (in set #23) Display or attach STH (e.g., to a wall) so it can be seen	They put up posters in supermarkets and distributed brochures. 彼らは、スーパーマーケットにポスターを貼り付け、パンフレットを配った。

	The gallery put up signs asking people to be careful around the art works. そのギャラリーは、芸術作品周辺に気を付けるように呼び掛ける標識を設置した。
50. move on (in set #23) Start doing or discussing STH new (job, activity, conversation topic)	This country can't move on until the president is elected. 大統領が選ばれるまで、その国は前へと進むことができない。
	Our conversation moved on to more congenial topics. 私たちの会話は、さらに気が合うような話題に移った。
85. throw out (in set #23) Refuse to accept or consider (esp. by people of authority)	His defense team tried to get the case thrown out. 彼の弁護団は、その訴訟を却下しようとした。
	A federal judge has thrown out a race and gender discrimination suit against the Community College of Philadelphia. 連邦裁判所判事が、フィラデルフィア・コミュニティ・カレッジに対して、人種差別と性差別における訴訟を取り下げた。
110. bring down (in set #23) Cause SB/STH to move downward or fall to the ground	She brought down the curtain on her career. 彼女は、自分のキャリアに幕を下ろした。
	The pilot managed to bring down the plane safely with great efforts. そのパイロットは、かなりの苦闘の結果、なんとか安全に飛行機を着陸させた。
6. find out (in set #24) Discover STH; obtain	She will find out the truth 10 years later. 彼女は 10 年後に、その真実を見つけ出すだろう。
	He learned how to skip school without his parents finding out. 彼は両親に気づかれずに、学校をさぼる方法を習得した。
45. bring up (in set #24) Raise for discussion or consideration	We're going to bring up the subject about the national budget on healthcare. 私たちは、医療における国家予算についての議題を提起するつもりだ。
	The study brings up the question as to what can be done to solve this situation. その研究では、この状況を解決するために何ができるのかについての疑問を提起している。
49. check out (in set #24) Have a look at; examine STH/SB (esp. to get more information or make a judgement)	For more information, check out the event's Facebook page. さらに詳しい情報については、イベントの Facebook ページをご覧ください。
	Please check out this video on YouTube if you want to know the event more. イベントについてもっと知りたい方は、YouTube 上のこの映像をご覧ください。

76. go through (in set #24) Experience STH difficult or unpleasant	The genius composer went through hell to compose a song for the victims. その天才作曲家は、被災者のために作曲しようと大変な苦労を経験した。
	I went through survival training as a military pilot. 私は軍事パイロットとして、サバイバル訓練を経験した。
120. move back (in set #24) Return to a place one has lived in before	Your family moved back to England when you were a teenager. あなたが 10 代の頃、家族はイングランドに戻った。
	She and her husband and their three children moved back to the ranch where they had lived. 彼女と夫と 3 人の子供たちは、以前生活していた牧場に戻った。

Narrative Appendix A: List of Phrasal Verbs

This appendix provides a list of phrasal verbs used in Experiment 3.

Week 1	Week 2	Week 3	Week 4	Week 5
set up	come up with	figure out	take off	take on
look down	look out	bring in	keep up	carry out
get down to	go through	turn up	hold out	turn over
write down	go around	start out	lay down	get in
sit back	set about	put off	give back	give out

Week 6	Week 7	Week 8	Week 9
work out	come out	take out	give up
bring back	go ahead	break down	take over
pull back	line up	turn back	shut down
stand out	turn down	make out	keep on
come through	get on with	sort out	pass on

Narrative Appendix B: Weekly Stories for Phrasal Verbs

This appendix includes stories that were turned into picture storyboards used in Experiment 3, with the target phrasal verbs underlined.

Week 1

a. Boy Scouts Camping: There are a large number of boy scouts on a camping trip. They have a long tradition of camping. When they arrive at their camp site, most kids set up their tents, while some take advantage of the trees to put up hammocks). After resting a bit they get down to chopping wood for a fire. They work hard.

b. Washing Windows: There is a man washing windows on a tall building; he looks down at the city; he is working hard to make money to enter college. Inside the building, he sees people sitting back in their chairs watching a soccer game. He is jealous. He has a strong desire to rest.

c. Lottery: A person is writing down numbers they see on the TV; They want to win the lottery because their mother is sick and has big bills. The whole family is watching the numbers. This is high drama! The number does not match, and they lose hope.

Week 2

a. Uber Driver: A person is trying to come up with a way to go through his life with a low income. He spends time doing research on second jobs. Then, he sets about his second job, being an Uber driver. With two jobs he will be able to improve the quality of his life.

b. Central Character Raven: The central figure in a story is looking out the window. He shouts so that people pay attention because a herd of goats is coming along the main road. Suddenly, ravens attack the goats! The person reading the story looks up because he hears ravens--there are many ravens outside his window in real life.

c. Having a Baby: A couple arrives at the hospital so she can have her baby. The man is going around the hospital, pacing with worry. His wife safely gives birth to a little baby. Today is a special occasion.

Week 3

a. New Business: A man is starting out building a new business. It is a fast-food restaurant called “Buddy’s”. He wants to do business in a part of town that used to have a lot of serious crime. However, in the past decade this area has slowly been getting safer. He reduces his risk of crime by installing an alarm system.

b. Zoo: There is a zoo in the city that is going to bring in a panda from China. Everyone loves pandas, with the possible exception of animal activists. The zoo puts off a renewal for a long time and public confidence in the way the animals are treated declines. Finally, the zoo builds bigger and more beautiful exhibits for the animals. Now everyone takes pride in their new zoo.

c. Missing Daughter: A couple are worried because their daughter has disappeared. They express concern that their daughter's husband might have tried to kill her. The police investigate and try to figure out what happened to her. Her body has not yet been found, so they take comfort in the possibility that she may still be alive. One day a month later, the searching turns up her in another city! The couple feel a deep sense of relief.

Week 4

a. Musician: The musician makes music. He sings about major concerns in the inner city. People try to keep up with his music. They are taking notice of him and he is becoming popular. Listeners feel his pain when they listen to his music. He holds a concert to give back to his city.

b. Dropped Earing: After taking off her clothes, a woman puts up her hand to her ear. She finds she has lost one earring. Her child finds the earring and holds out his

hand to give it to her. But she finds it is broken! This is a big surprise. The woman is unhappy.

c. Exercise: A woman starts to exercise to lose weight and build muscle. She takes pleasure in her strong body. After exercising with a dumbbell, she lays it down and lies in her bed. She sleeps to save energy for tomorrow's work.

Week 5

a. Bank Robbery Getaway: Two men are carrying out bank robbery. The thief is out motioning for his partner to get in the getaway car. The partner is bringing out a heavy bag of money from the bank. The thieves start driving and soon a police car follows them. Now a chase is taking place! The thieves try to change direction suddenly and when the partner takes a quick glance behind them, it seems they are free! At night the thieves are at home watching television with their bag of money. They got away with their crime.

b. Homeless Fisherman: A homeless man is in a difficult situation. He is cold and has no money. He gives out cards that say "Help me. I will work for food." A man shows him how to catch fish. He begins to work hard, make money and make progress by buying new clothes and renting an apartment. In the end, he takes on an employee in his fish selling business.

c. Soap Box Dog Killer: There is a person standing on a box in the park. Some people are listening to him while others are booing him. He scolds the booing people. There is free speech in this country! He can talk! He says he just wants to give voice to the people who think all dogs should be killed. They are dangerous! He says he has a good reason for wanting to kill all the dogs--a pack of dogs attacked him in the park. Now he is turning over his place at the box to a different speaker. She thinks they should kill all the cats!

Week 6

a. Senior Artist: An old man, a senior citizen, is painting a picture. It seems he is an artist because there are many pictures lying around. He is outside in a field and he is painting the features of the landscape around him. It appears his favorite subject to paint is landscapes. The natural beauty of the area comes through in his painting. The mountain in the background especially stands out. Suddenly, a heavy rain begins and the man picks up his pictures and hurries to find shelter.

b. Hunting: A father and his son are walking through the woods. They are going hunting. They are making small talk, just chatting about the weather. The young man feels pressure to kill something today as he was not able to hit anything before. He wants to bring back at least 2 ducks for his mother to cook. The father and son

see a big flock of birds on a lake. The son pulls back the trigger on his gun. The birds take flight and the young man is able to shoot one down.

c. Black Lives Matter: A group is protesting police violence (Black Lives Matter). They take issue with more violence against black people than white people by police. Police violence has a negative impact on black people more than white. People hope they can work out a situation where police will be fair to everyone and rules about fair treatment can become law.

Week 7

a. Model Man: A woman runs into an elevator, but she is bumped into by a man. He says to her 'go ahead.' She feels that he is a gentleman. She squeezes into the one open space in the elevator, so the man does not get on the elevator. She gets on with her work until 6 p.m. After finishing her work, she goes into a bookstore to get a new magazine and she finds the man who bumped into her is a model on the hard copy of a magazine cover.

b. Cake Shop: A woman takes note of what she needs to buy today and goes shopping. She has heard on mass media that there is a new cake shop opening today. She doesn't know where it is, but walking down the street, she sees a cake shop that smells sweet and people are lining up to get cakes. After she eats one of their cakes,

the long line makes sense. The shop is very popular because their food is great. There are a wide range of cakes. She is glad to have the cake shop in her neighborhood.

c. Waterfall Propose: A couple are on the side of the waterfall. The man kneels down to propose to the woman when a rainbow comes out. He thinks “this is a sign that we are a perfect match”. But the woman turns down his proposal because she has another boyfriend. He tries to be cheerful, but it is human nature that he gets depressed after this. He loses sight of his goals and becomes fat and lazy.

Week 8

a. Serial Killer: A serial killer tries to commit murder and sneak into a house, but he makes noise and the residents notice there is someone in their house. The father takes out a gun and points it at the murderer, who runs away. The family feels afraid that he might turn back and kill them later.

b. Preparation for Storm: There is a house by the ocean. A storm with a dark cloud is coming close. Most people stay at home to wait for the storm to pass. The power lines break down from the high winds. The wind is loud and frightening but in the middle of the hurricane everything becomes quiet. In the eye of the storm there is a

long silence, then the wind starts howling again. Actually, the general population prepared for the storm, but a few people did not have any protection.

c. Monitoring Camera Technology: In China there is a top-secret plan to catch people who commit a crime. Scientists are using technology to make out who people are. The technology is important feature to have public order and security. People can be identified after just 2 seconds because of all the cameras on the streets. The government has a high degree of confidence that they can gain knowledge about any citizen. The government wants to sort out any problems caused by criminals and put them in jail.

Week 9

a. Pottery Maker: A master pottery maker is working at his kiln. He has an apprentice. The master hopes to pass on his skills to his apprentice. In the long term, he hopes that when he retires this boy will be able to take over his successful business. One basic rule of the master is to always keep the working area clean. The other rule is to always have customers pay cash. He does not accept credit cards. The pottery maker has had the good fortune of being popular with tourists.

b. Prime Minister's Duty: The prime minister is worried about the financial situation of her country. Because of COVID-19 there has been a significant decrease in taxes

this year. Taxes are an important component of running a country--they help pay for the police, for roads, and transportation. The prime minister is also worried about businesses in the private sector as well. Many restaurants and bars have shut down. Without government help the remaining business owners may give up. The prime minister will give money to private businesses as well as cities.

c. Injured Solider: There are soldiers walking through the snow, carrying their guns. They keep on marching even though they are cold and tired. Suddenly, one of the soldiers steps on a landmine. A nurse from the red cross takes care of him. The soldier hands the nurse a letter to send word to his family that he is still alive.

Week 10 (No target phrasal verbs in week 10)

a. Earthquake: A family is watching TV all together when suddenly the room starts shaking. It is an earthquake! The family takes refuge from falling items by hiding under their kitchen table. Many people have lost their entire homes in this natural disaster. Groups of aid workers are providing relief to people by handing out blankets and water bottles.

b. Soccer Match: Fans are watching a soccer match and rooting for their team. With the score 3-0 though, many fans have lost faith that their team will win and seem very depressed. The fans at home feel sympathy for those fans in the stadium. They

also feel sad when they see the depressed fans on TV. The other team declares victory with the final score 5-0. It's a blowout! The other team's mascot, a tiger, is menacing the losing team's mascot, a bee.

c. Mermaid Ball: The scene is the mermaid ball. Mermaids who have crowns on their head are given priority and get to go in the fancier door, while mermaids without crowns gain admission only by going to the back door. The mermaid royalty (the ones with crowns) have access to a free bar. The mermaids without crowns have to buy their food from the vending machines. There are the mermaid King and his son the prince at the ball. The prince is taking a poll and collecting data from the mermaids in order to decide who to marry. That's a weird way to find a marriage partner!

Narrative Appendix C: Phrasal Verb Test Items for Gap-filling Tests

This appendix contains sentences and their corresponding Japanese translations for 45 phrasal verbs that were used in the phrasal verb gap-filling tests of Experiment 3. For the pre-test, English sentence for each phrasal verb was obtained from the PHaVE List (Garnier and Davies, 2014), while for the post-test, sentences were drawn from Japanese-English or English-English online dictionaries such as, *Cambridge Dictionary*, *Collins English Dictionary*, *Oxford Learner's Dictionaries*, *Longman Dictionary of Contemporary English*, and *Alc*. Some of sentences were slightly changed to make them easier to understand. Any Japanese translations for sentences that had not been previously translated were provided by the author.

Week	Phrasal verb	Test	Sentences
Week 1	set up	pre	An advisory committee is being set up. 諮問委員会が現在設置されている。
		post	Sometimes schools set up a meeting with the parents concerning their children. 学校は時々、子供たちについて両親と面談を設けることがある。
	look down	pre	She looked down at the ground to see what she stepped on. 彼女は何を踏んだのか確認しようと、地面を見下ろした。
		post	He looked down at his hands and he didn't reply. 彼は自分の手を見下ろして、返事をしなかった。
	get down to	pre	We should get down to discussing those issues as soon as possible. できるだけ早くそれらの問題について議論に取りかかる必要がある。
		post	New York City's new mayor got down to business, but there was a battle over his plan. ニューヨークの新市長は事業に真剣に取りかかったが、彼の計画をめぐってひと悶着あった。

	write down	pre	You should write down his contact details in case you want to get in touch. 連絡を取りたい時のために、彼の連絡先の詳細を記録しておくべきだ。
		post	The detective wrote down the names of all his brothers and sisters. その探偵は、彼のすべての兄弟姉妹の名前を記録した。
	sit back	pre	She sat back in her chair and turned on the TV. 彼女は椅子に深く座り、テレビの電源を入れた。
		post	The pilot announced “we will be taking off shortly.” I sat back in my seat and fastened my seat belt. 機長の「間もなく離陸する」というアナウンスを聞き、私は座席にゆったりと座り、シートベルトを締めた。
Week 2	come up with	pre	She instantly came up with a solution to the problem. 彼女はすぐにその問題の解決策を思いついた。
		post	I have thought through every possibility, but I couldn’t come up with any other solution. あらゆる可能性を検討してみたが、他の解決策を思いつくことができなかった。
	look out	pre	She liked to go by the window and look out at the garden. 彼女は窓際に行き、窓からその庭を眺めるのが好きだった。
		post	The posters feature a woman looking out the back window of a van. そのポスターは、バンの後ろの窓から外を見る女性が特徴的である。
	go through	pre	You have to understand the tough situation she went through before judging her. 彼女について判断する前に、彼女が経験した厳しい状況を理解してあげないと。
		post	Karen’s husband had gone through a period of unemployment. カレンの夫は、失業期間を経験したことがある。
	go around	pre	There is a rumor going around that she is pregnant. 彼女が妊娠しているという噂が広まっている。
		post	Plenty of food was served and the bride and groom went around the tables to greet the guests. たくさんの食事が出され、新郎新婦はテーブルを一回りして招待客に挨拶した。
	set about	pre	We set about laying the table before our guests arrived. 私たちはお客様が到着する前に、テーブルの配置に取り掛かった。
		post	He set about decorating their new house. 彼は新居の内装に取りかかった。
Week 3	figure out	pre	Despite her efforts, she couldn’t figure out what had happened. 努力のかいもなく、彼女は何が起こったのか理解できなかった。

		post	It's crucial that you take a step back and figure out what's most important to you. 一歩引いて、あなたにとって何が最も重要かを理解することは大切だ。
	bring in	pre	I brought in my laptop computer today because my office computer is broken. 自分のオフィスのパソコンが壊れているので、今日は自分のパソコンを持ち込んだ。
		post	We generally ask patients not to bring in foods. 基本的に患者さんには、食べ物の持ち込みはお断りしております。
	turn up	pre	The search turned up solid evidence against him. その捜査は、彼に対する確かな証拠を見つけ出した。
		post	The examination of Mr. Martin's computers has so far turned up no evidence that it was hacked. マーティンさんのコンピュータの調査では、今のところハッキングされた証拠は見つけ出されていない。
	start out	pre	She started out as a shop assistant and gradually climbed the employment ladder. 彼女は店員として仕事を始め、徐々に昇進の階段を昇っていった。
		post	She started out as a model and later turned to acting. 彼女はモデル出身で後に女優を始めた。
	put off	pre	Now that I had more free time, there was no excuse to put off exercising any longer. 余暇が増えたので、運動を後回しにする言い訳はもうなかった。
		post	Nearly a third of people between 18 and 34 have put off marriage or having a baby due to the recession. 18 歳から 34 歳までの約 3 分の 1 の人が、不況のために結婚や出産を先延ばしにしている。
Week 4	take off	pre	I took off my shirt and went to bed. 私はシャツを脱いで寝た。
		post	He took off his clothes and got in the shower. 彼は服を抜いで、シャワーを浴びた。
	keep up	pre	Workers' income has not kept up with inflation. 労働者の収入はインフレに追いつけていない。
		post	We have to create 150,000 jobs every month just to keep up with population growth. 人口増加に同じペースで追いつくために、毎月 15 万人分の雇用を創出しなければならない。
	hold out	pre	He took the keys and held them out to her. 彼はカギを取って、彼女に差し出した。
		post	"Do you know what these things are called?" the boy asked, holding out his palm with the ant on it.

			「これなんて呼ぶか知ってる？」と、アリを乗せた手のひらを差し出して、少年が聞いてきた。
	lay down	pre	I laid down my book and stood up. 私は本を置いて、立ち上がった。
		post	Silvia surrendered and laid down her weapons. シルビアは今降伏し、武器を置いた。
	give back	pre	It's nice to be able to give back to the community. 地域に還元できることは、素晴らしいことです。
		post	The baseball player said he wanted to give something back to the people of Los Angeles after the support he received when he was injured. その野球選手は、けがをした時に声援をもらったので、ロスの人々に恩返しをしたいと語った。
Week 5	take on	pre	Nobody was willing to take on such an awful job. 誰もそのような大変な仕事なんて引き受けようとはしなかった。
		post	He is now taking on this political role because it's expedient. 急場しのぎのために、彼は現在政治的にこの役割を引き受けている。
	carry out	pre	The experiment was carried out by a well-known academic. その実験は有名な学者によって行われた。
		post	The bomb attacks were carried out by terrorists. その爆弾による攻撃は、テロリストによって実行された。
	turn over	pre	The police officer turned over the criminal to the jail guard. 警官はその犯人を刑務所の看守に引き渡した。
		post	The information was turned over to the special prosecutor. その情報は特別検察官に引き渡された。
	get in	pre	The new security lock prevents thieves from getting in. 新しいセキュリティロックは、泥棒の侵入を防ぐ。
		post	She walked to my car, got in, and drove away. 彼女は車の方へ歩き、中に入り(車に乗り)、去っていった。
	give out	pre	The committee gave out more than 100 copies in the last meeting. その委員会は、前回の会議で 100 部以上のコピーを配布した。
		post	Students were giving out leaflets to everyone on the street. 学生たちは道行く人にビラを配布していた。
Week 6	work out	pre	We still need to work out the details of the procedure. 私たちはまだその進行の細かい部分を計画する必要があります。
		post	The details will be worked out in subsequent meetings. 詳細は次のミーティングで計画されるだろう。
	bring back	pre	This will bring back war into the country. これによって、その国に戦争を呼び戻すでしょう。
		post	I was there during the war. So the movie brings back a lot of images of that time.

			戦時中そこにいたので、その映画はたくさんの戦争のイメージを呼び戻す。
	pull back	pre	She pulled back the hair from her face. 彼女は、顔の前にあった髪を退けた。
		post	She pulled back the curtain to see the sunrise. 日の出を見るために、カーテンを退けた。
	stand out	pre	Excellent product quality is what made the brand stand out from its competitors. 優れた製品品質は、そのブランドを競合会社から際立たせている。
		post	The girl was tall, with long dark hair; she stood out in the crowd. その少女は身長が高く、人ごみの中でも長い黒髪が際立っていた。
	come through	pre	Her disappointment came through by the tone of her voice. 彼女の落ち込んだ感情が、声の調子によってはっきりと表れていた。
		post	Your passion for truth, your determination, your integrity--it all comes through in the book. あなたの真実への情熱、決意、誠実さ。そのすべてがこの本にはっきりと表れている。
Week 7	come out	pre	She went into the bank and came out with some money. 彼女は銀行へ行き、いくらかお金を持って出てきた。
		post	People are always surprised to hear such a big voice coming out of such a little kid. その小さな子供から出てくる大きな声を聞いて、いつもみんな驚く。
	go ahead	pre	Go ahead and ask me your question! さあどうぞ、質問してください。
		post	A: Let me ask you a question. B: Go ahead. A: 質問させてください。 B: どうぞ続けて。
	line up	pre	Dozens of taxis were lined up at the entrance. 入口には、たくさんのタクシーが並んでいた。
		post	The kids always line up to wash their hands before dinner. 子供たちはいつも、夕食前に手を洗うため並んでいる。
	turn down	pre	This is an opportunity you would be foolish to turn down. この機会を断るのは愚かだ。
		post	The singer has never turned down a request for an autograph or a picture from her fans. その歌手は、ファンからのサインや写真のお願いを決して断ったことがない。
	get on with	pre	We might as well get on with it if we want to finish on time. 時間通りに終わらせたいのなら、それをやり続けた方が良い。
		post	Be quiet and get on with your work. 静かにして、仕事を続けてください。

Week 8	take out	pre	He tore open the envelope and took out a few bills. 彼は封筒を破って、何枚かの請求書を取り出した。
		post	He took out a handkerchief and swiped at the sweat beading on his forehead. 彼はハンカチを取り出して、額に汗の水滴を拭いた。
	break down	pre	Our car broke down yesterday. 私たちの車は、昨日壊れた。
		post	Someone broke down my door and got into my apartment. 誰かがドアを壊して、私のアパートに入ってきた。
	turn back	pre	She turned around and walked to the door. 彼女は振り返り、ドアへと歩いて行った。
		post	There's no way we can turn back on this road. この道を引き返すのは無理だ。
	make out	pre	I could barely make out his face in the dark. 私は、暗闇の中でも彼の顔をかろうじて識別することができた。
		post	The light was insufficient to make out the features in the painted faces in the portraits that lined the walls. 壁に並んでいる肖像画の描かれた顔の特徴を識別するのに、光が足りなかった。
	sort out	pre	A few ideas were raised to sort out the company's financial issues. 会社の財務問題を解決するため、いくつかのアイデアが挙げられた。
		post	She went to a psychiatrist to try to sort out her problems. 彼女は、自分の問題を解決しようと、精神科医を訪れた。
Week 9	give up	pre	She had to give up smoking when she got pregnant. 彼女は妊娠した時、喫煙をやめなければならなかった。
		post	When I told him about my dream, he said, "You only get one life. Don't give up your dream just because you are blind." 私が自分の夢について先生に話すと、先生は、「人生は一度きり。目が不自由だからというだけで夢を諦めてはいけない」とおっしゃいました。
	take over	pre	After her father died, she took over the company. 父が亡くなった後、彼女はその会社を引き継いだ。
		post	He reluctantly took over the family business after the death of his father. 彼は父の死後、家業を仕方なく引き継ぎました。
	shut down	pre	You should shut down your computer at night to save electricity. 電気を節約するには、夜間にコンピュータを停止させるべきだね。
		post	A lot of companies are shutting down non-essential operations to save energy. 多くの会社は、節約のために不必要な機械の運転を停止している。
	keep on	pre	She wiped tears off her cheeks but kept on crying.

			彼女は頬の涙をぬぐったが、泣き続けた。
		post	Can the business keep on expanding without hiring more people? 更に雇用せずに、事業は止むことなく拡大し続けられるのだろうか？
	pass on	pre	I got this message this morning and was asked to pass it on to you. 私は今朝このメッセージを受け取り、あなた伝えるように頼まれました。
		post	The city tries hard to pass on information to residents about utility savings programs. その市は、光熱費節約プログラムの情報を住民に伝える努力をしている。

Narrative Appendix D: Stories in Pre-test and Post-test

This appendix includes stories that were used in the pre-test and post-test of Experiment 3. Phrasal verbs that participants studied are underlined.

a. Bank Robbery: A very serious crime has happened at the bank. A robber took advantage of an open safe and tried to steal a lot of money. However, the police arrived and the robber faced a difficult situation. He <u>came up with</u> a plan to escape the police. He took a hostage! He tried to <u>carry out</u> his plan by threatening to shoot the hostage while he walked out of the bank. One of the policemen sneezed and accidentally shot his gun, so the robber committed murder and killed his hostage. The police immediately shot the robber so two people died in the attempted bank robbery.

b. Black's Gyms: Lately a man named Erik Black has been in the news. He is the owner of a chain of gyms where men only can work out and build muscles. Erik Black looks very strong. He says today's men have lost sight of what it means to be a man. He says as a basic rule, men are stronger than women, and they need to show it. They need muscles. Erik's chain of gyms <u>brings in</u> many men who have a strong desire to get very muscular. It's hard to <u>sort out</u> whether Erik's gyms discriminate against women.

c. Special Date: A couple are dressed up for a special occasion. Then just the man enters a house. He sets up a table with a cake and flowers. Later other people are waiting inside and are lined up to celebrate some event. When the woman comes in everyone throws confetti and releases balloons. It is a big surprise! The man gives the woman a big bouquet of flowers. They smell sweet. Later everyone starts clapping and yelling “Kiss! Kiss!” and the couple feel pressure to kiss. So they do.

Narrative Appendix E: List of Collocations

This appendix provides a list of collocations used in Experiment 4.

Week 1	Week 2	Week 3	Week 4	Week 5
high drama	low income	fast food	save energy	difficult situation
large number	improve quality	possible exception	inner city	catch fish
work hard	main road	past decade	major concern	quick glance
long tradition	special occasion	public confidence	big surprise	change direction
strong desire	central figure	serious crime	take pleasure	free speech
make money	spend time	take pride	feel pain	take place
lose hope	give birth	express concern	lose weight	give voice
enter college	real life	take comfort	take notice	watch television
take advantage	pay attention	do business	make music	make progress
whole family	do research	reduce risk	build muscle	good reason

Week 6	Week 7	Week 8	Week 9	Week 10
become law	smell sweet	high degree	important component	natural disaster
favorite subject	human nature	important feature	significant decrease	declare victory
small talk	open space	dark cloud	red cross	provide relief
natural beauty	wide range	top secret	private sector	have access
senior citizen	perfect match	gain knowledge	send word	lose faith
negative impact	lose sight of	make noise	prime minister	take refuge
heavy rain	take note	commit murder	good fortune	gain admission
take issue	make sense	long silence	pay cash	feel sympathy
take flight	hard copy	general population	basic rule	collect data
feel pressure	mass media	use technology	long term	give priority

Narrative Appendix F: Weekly Stories for Collocations

This appendix includes stories that were turned into picture storyboards used in Experiment 4, with the target collocations underlined.

Week 1

a. Boy Scouts Camping: There are a large number of boy scouts on a camping trip. They have a long tradition of camping. When they arrive at their camp site, most kids set up their tents, while some take advantage of the trees to put up hammocks). After resting a bit they get down to chopping wood for a fire. They work hard.

b. Washing Windows: There is a man washing windows on a tall building; he looks down at the city; he is working hard to make money to enter college. Inside the building, he sees people sitting back in their chairs watching a soccer game. He is jealous. He has a strong desire to rest.

c. Lottery: A person is writing down numbers they see on the TV; They want to win the lottery because their mother is sick and has big bills. The whole family is watching the numbers. This is high drama! The number does not match, and they lose hope.

Week 2

a. Uber Driver: A person is trying to come up with a way to go through his life with a low income. He spends time doing research on second jobs. Then, he sets about his second job, being an Uber driver. With two jobs he will be able to improve the quality of his life.

b. Central Character Raven: The central figure in a story is looking out the window. He shouts so that people pay attention because a herd of goats is coming along the main road. Suddenly, ravens attack the goats! The person reading the story looks up because he hears ravens--there are many ravens outside his window in real life.

c. Having a Baby: A couple arrives at the hospital so she can have her baby. The man is going around the hospital, pacing with worry. His wife safely gives birth to a little baby. Today is a special occasion.

Week 3

a. New Business: A man is starting out building a new business. It is a fast-food restaurant called "Buddy's". He wants to do business in a part of town that used to have a lot of serious crime. However, in the past decade this area has slowly been getting safer. He reduces his risk of crime by installing an alarm system.

b. Zoo: There is a zoo in the city that is going to bring in a panda from China. Everyone loves pandas, with the possible exception of animal activists. The zoo puts off a renewal for a long time and public confidence in the way the animals are treated declines. Finally, the zoo builds bigger and more beautiful exhibits for the animals. Now everyone takes pride in their new zoo.

c. Missing Daughter: A couple are worried because their daughter has disappeared. They express concern that their daughter's husband might have tried to kill her. The police investigate and try to figure out what happened to her. Her body has not yet been found, so they take comfort in the possibility that she may still be alive. One day a month later, the searching turns up her in another city! The couple feel a deep sense of relief.

Week 4

a. Musician: The musician makes music. He sings about major concerns in the inner city. People try to keep up with his music. They are taking notice of him and he is becoming popular. Listeners feel his pain when they listen to his music. He holds a concert to give back to his city.

b. Dropped Earing: After taking off her clothes, a woman puts up her hand to her ear. She finds she has lost one earring. Her child finds the earring and holds out his

hand to give it to her. But she finds it is broken! This is a big surprise. The woman is unhappy.

c. Exercise: A woman starts to exercise to lose weight and build muscle. She takes pleasure in her strong body. After exercising with a dumbbell, she lays it down and lies in her bed. She sleeps to save energy for tomorrow's work.

Week 5

a. Bank Robbery Getaway: Two men are carrying out bank robbery. The thief is out motioning for his partner to get in the getaway car. The partner is bringing out a heavy bag of money from the bank. The thieves start driving and soon a police car follows them. Now a chase is taking place! The thieves try to change direction suddenly and when the partner takes a quick glance behind them, it seems they are free! At night the thieves are at home watching television with their bag of money. They got away with their crime.

b. Homeless Fisherman: A homeless man is in a difficult situation. He is cold and has no money. He gives out cards that say "Help me. I will work for food." A man shows him how to catch fish. He begins to work hard, make money and make progress by buying new clothes and renting an apartment. In the end, he takes on an employee in his fish selling business.

c. Soap Box Dog Killer: There is a person standing on a box in the park. Some people are listening to him while others are booing him. He scolds the booing people. There is free speech in this country! He can talk! He says he just wants to give voice to the people who think all dogs should be killed. They are dangerous! He says he has a good reason for wanting to kill all the dogs--a pack of dogs attacked him in the park. Now he is turning over his place at the box to a different speaker. She thinks they should kill all the cats!

Week 6

a. Senior Artist: An old man, a senior citizen, is painting a picture. It seems he is an artist because there are many pictures lying around. He is outside in a field and he is painting the features of the landscape around him. It appears his favorite subject to paint is landscapes. The natural beauty of the area comes through in his painting. The mountain in the background especially stands out. Suddenly, a heavy rain begins and the man picks up his pictures and hurries to find shelter.

b. Hunting: A father and his son are walking through the woods. They are going hunting. They are making small talk, just chatting about the weather. The young man feels pressure to kill something today as he was not able to hit anything before. He wants to bring back at least 2 ducks for his mother to cook. The father and son

see a big flock of birds on a lake. The son pulls back the trigger on his gun. The birds take flight and the young man is able to shoot one down.

c. Black Lives Matter: A group is protesting police violence (Black Lives Matter). They take issue with more violence against black people than white people by police. Police violence has a negative impact on black people more than white. People hope they can work out a situation where police will be fair to everyone and rules about fair treatment can become law.

Week 7

a. Model Man: A woman runs into an elevator, but she is bumped into by a man. He says to her 'go ahead.' She feels that he is a gentleman. She squeezes into the one open space in the elevator, so the man does not get on the elevator. She gets on with her work until 6 p.m. After finishing her work, she goes into a bookstore to get a new magazine and she finds the man who bumped into her is a model on the hard copy of a magazine cover.

b. Cake Shop: A woman takes note of what she needs to buy today and goes shopping. She has heard on mass media that there is a new cake shop opening today. She doesn't know where it is, but walking down the street, she sees a cake shop that smells sweet and people are lining up to get cakes. After she eats one of their cakes,

the long line makes sense. The shop is very popular because their food is great. There are a wide range of cakes. She is glad to have the cake shop in her neighborhood.

c. Waterfall Propose: A couple are on the side of the waterfall. The man kneels down to propose to the woman when a rainbow comes out. He thinks "this is a sign that we are a perfect match". But the woman turns down his proposal because she has another boyfriend. He tries to be cheerful, but it is human nature that he gets depressed after this. He loses sight of his goals and becomes fat and lazy.

Week 8

a. Serial Killer: A serial killer tries to commit murder and sneak into a house, but he makes noise and the residents notice there is someone in their house. The father takes out a gun and points it at the murderer, who runs away. The family feels afraid that he might turn back and kill them later.

b. Preparation for Storm: There is a house by the ocean. A storm with a dark cloud is coming close. Most people stay at home to wait for the storm to pass. The power lines break down from the high winds. The wind is loud and frightening but in the middle of the hurricane everything becomes quiet. In the eye of the storm there is a

long silence, then the wind starts howling again. Actually, the general population prepared for the storm, but a few people did not have any protection.

c. Monitoring Camera Technology: In China there is a top-secret plan to catch people who commit a crime. Scientists are using technology to make out who people are. The technology is important feature to have public order and security. People can be identified after just 2 seconds because of all the cameras on the streets. The government has a high degree of confidence that they can gain knowledge about any citizen. The government wants to sort out any problems caused by criminals and put them in jail.

Week 9

a. Pottery Maker: A master pottery maker is working at his kiln. He has an apprentice. The master hopes to pass on his skills to his apprentice. In the long term, he hopes that when he retires this boy will be able to take over his successful business. One basic rule of the master is to always keep the working area clean. The other rule is to always have customers pay cash. He does not accept credit cards. The pottery maker has had the good fortune of being popular with tourists.

b. Prime Minister's Duty: The prime minister is worried about the financial situation of her country. Because of COVID-19 there has been a significant decrease in taxes

this year. Taxes are an important component of running a country--they help pay for the police, for roads, and transportation. The prime minister is also worried about businesses in the private sector as well. Many restaurants and bars have shut down. Without government help the remaining business owners may give up. The prime minister will give money to private businesses as well as cities.

c. Injured Solider: There are soldiers walking through the snow, carrying their guns. They keep on marching even though they are cold and tired. Suddenly, one of the soldiers steps on a landmine. A nurse from the red cross takes care of him. The soldier hands the nurse a letter to send word to his family that he is still alive.

Week 10

a. Earthquake: A family is watching TV all together when suddenly the room starts shaking. It is an earthquake! The family takes refuge from falling items by hiding under their kitchen table. Many people have lost their entire homes in this natural disaster. Groups of aid workers are providing relief to people by handing out blankets and water bottles.

b. Soccer Match: Fans are watching a soccer match and rooting for their team. With the score 3-0 though, many fans have lost faith that their team will win and seem very depressed. The fans at home feel sympathy for those fans in the stadium. They

also feel sad when they see the depressed fans on TV. The other team declares victory with the final score 5-0. It's a blowout! The other team's mascot, a tiger, is menacing the losing team's mascot, a bee.

c. Mermaid Ball: The scene is the mermaid ball. Mermaids who have crowns on their head are given priority and get to go in the fancier door, while mermaids without crowns gain admission only by going to the back door. The mermaid royalty (the ones with crowns) have access to a free bar. The mermaids without crowns have to buy their food from the vending machines. There are the mermaid King and his son the prince at the ball. The prince is taking a poll and collecting data from the mermaids in order to decide who to marry. That's a weird way to find a marriage partner!

Narrative Appendix G: Stories in Pre-test and Post-test

This appendix includes stories that were used in the pre-test and post-test of Experiment 4. Collocations that participants studied are underlined.

a. Bank Robbery: A very serious crime has happened at the bank. A robber took advantage of an open safe and tried to steal a lot of money. However, the police arrived and the robber faced a difficult situation. He came up with a plan to escape the police. He took a hostage! He tried to carry out his plan by threatening to shoot the hostage while he walked out of the bank. One of the policemen sneezed and accidentally shot his gun, so the robber committed murder and killed his hostage. The police immediately shot the robber so two people died in the attempted bank robbery.

b. Black's Gyms: Lately a man named Erik Black has been in the news. He is the owner of a chain of gyms where men only can work out and build muscles. Erik Black looks very strong. He says today's men have lost sight of what it means to be a man. He says as a basic rule, men are stronger than women, and they need to show it. They need muscles. Erik's chain of gyms brings in many men who have a strong desire to get very muscular. It's hard to sort out whether Erik's gyms discriminate against women.

c. Special Date: A couple are dressed up for a special occasion. Then just the man enters a house. He sets up a table with a cake and flowers. Later other people are waiting inside and are lined up to celebrate some event. When the woman comes in everyone throws confetti and releases balloons. It is a big surprise! The man gives the woman a big bouquet of flowers. They smell sweet. Later everyone starts clapping and yelling “Kiss! Kiss!” and the couple feel pressure to kiss. So they do.