Korean Journal of English Language and Linguistics, Vol 26, January 2026, pp. 1-20
DOI: 10.15738/kjell.26..202601.1

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

Comparing Form-focused Retrieval and Meaning-inferencing Approaches to Phrasal Verb Learning by Korean EFL Learners*

Boha Hong · Hyunsook Yoon (Hankuk University of Foreign Studies)

Received: December 20, 2025
Revised: January 13, 2026
Accepted: January 13, 2026

Hong, Boha (First author)
Department of TESOL, Graduate School, Hankuk University of Foreign Studies
107 Imun-ro, Dongdaemun-gu
Seoul, Korea
Tel: +82-2-2173-3978
Email: selenaboha@gmail.com

Yoon, Hyunsook (Corresponding author)
Department of English Education
Hankuk University of Foreign Studies
107 Imun-ro, Dongdaemun-gu
Seoul, Korea
Tel: +82-2-2173-3978
Email: hsyoon3@hufs.ac.kr

* This work was supported by Hankuk University of Foreign Studies Research Fund of 2025.

ABSTRACT

Hong, Boha and Hyunsook Yoon. 2026. Comparing form-focused retrieval and meaning-inferencing approaches to phrasal verb learning by Korean EFL learners. *Korean Journal of English Language and Linguistics 26*, 1-20.

English phrasal verbs (PVs) pose significant learning challenges for EFL learners due to their structural complexity and non-compositional meaning. Yet few studies have directly compared form-focused and meaning-focused approaches to PV learning. To address this gap, the present study examined the effects of two form-focused methods, particle-focused retrieval (PR) and chunk-focused retrieval (CR), and one meaning-focused method, meaning-inferencing (MI), conceptualized as structural and semantic elaboration, respectively. Using a within-subject design, 36 Korean adult EFL learners completed all three conditions. Learners' recall of both PV form and meaning was assessed through immediate and two-week delayed post-tests, supplemented by surveys and semi-structured interviews. Results revealed no significant differences among PR, CR, and MI for meaning recall. For form recall, PR outperformed the other conditions at both time points, although its advantage diminished over time and became comparable to CR. Learner perceptions did not align with performance outcomes: MI was most preferred despite weaker recall, whereas PR, the least favored, yielded the strongest results. These findings underscore the effectiveness of retrieval-based practice, particularly particle-focused retrieval, for supporting accurate PV form learning, while suggesting that meaning-inferencing strategies may be more sensitive to individual learner characteristics. The study contributes to a more nuanced understanding of how different learning conditions support distinct aspects of complex multi-word expression learning.

KEYWORDS

phrasal verbs, retrieval practice, form-focused learning, meaning inferencing, multi-word expressions

1. Introduction

Learning L2 vocabulary begins with knowing its form and meaning (Jiang 2002, Schmitt and Schmitt 2020), which constitute the most fundamental components within Nation's (2013) comprehensive taxonomy of word knowledge. Acquisition depends on where attention is placed (Barcroft 2002), which reflects the principle that "we learn what we focus on" (Nation 2024, p. 3). It is essential, therefore, to identify effective instructional approaches for targeting specific lexical aspects in order to support vocabulary development by directing learners' attention either towards form or meaning (Boers et al. 2017, Nation 2024). Barcroft (2002) distinguished two such approaches: structural elaboration, which emphasizes formal properties, such as spelling, pronunciation, or construction, and semantic elaboration, which engages with an item's deeper conceptual comprehension beyond surface characteristics (Candry et al. 2017).

One well-established method of structural elaboration is retrieval practice, a conscious act of recalling information (Roediger and Karpicke 2006), for which considerable evidence demonstrates benefits for long-term memory, as shown in the "desirable difficulties" framework (Bjork and Kroll 2015). Recent studies on the structural complexity of multi-word units (MWUs)—recurring combinations of a particular word and the words that accompany it, as opposed to single words (Nation 2013)—underscore the need to extend retrieval-based methods beyond single-word form learning. Among MWUs, English phrasal verbs (PVs) are known to be both highly prevalent in language use and notoriously difficult for L2 learners (Boers 2014, Garnier and Schmitt 2016). In the verb + particle combinations of PVs (e.g., *give up* and *take off*), particles pose major obstacles because they are semantically elusive and can dramatically change a verb's meaning (Strong and Boers 2019b). As such, much research has centered on particles and has emphasized the importance of particle-focused retrieval practice.

However, since PVs convey non-compositional meanings that cannot easily be inferred from their constituent parts, they should also be treated as holistic units to capture idiomaticity. Applied psycholinguistics has supported the cognitive benefits of storing and processing MWUs holistically to enhance fluency (Ellis 2012, Wray 2002), and this view is further supported by evidence from eye-tracking and read-aloud studies (Ellis et al. 2008). Still, empirical research on holistic PV learning in EFL contexts remains largely underexplored, making it a promising area that warrants further study.

In contrast, this holistic approach has been widely examined in research on collocation, another key category of MWUs (Lewis 1993). Since collocations (e.g., pay attention) and PVs share structural similarities, with two words functioning as a single unit, collocation studies may provide useful insights for PV instruction. In particular, a substantial body of research on collocation learning has systematically compared chunk-focused and part-focused instructional approaches, demonstrating differential effects on form recall and learning efficiency (e.g., Boers et al. 2006, 2014, Ferguson et al. 2024). However, despite the structural similarity between collocations and phrasal verbs, such comparative investigations have rarely been extended to PV learning. This gap leaves unanswered the question of whether particle-focused retrieval (PR) and chunk-focused retrieval (CR)—both grounded in prior collocation research—yield distinct learning outcomes when applied to phrasal verbs.

Beyond structural elaboration, semantic elaboration represents another important dimension of vocabulary learning. Semantic elaboration involves contextualizing lexical items and linking new information to existing knowledge, a process that is particularly relevant for PVs given their opaque and context-dependent meanings (Pellicer-Sánchez 2020, Peters 2014). One of the key methods of semantic elaboration is meaning inferencing (MI), in which learners infer the meaning of unfamiliar expressions from contextual cues without immediately consulting a dictionary. Because the meanings of many PVs cannot be reliably derived from their component parts, MI may offer a complementary learning route that differs fundamentally from form-focused retrieval approaches, relying more heavily on discourse-level processing than on explicit attention to lexical form.

The effectiveness of MI is grounded in the Levels of Processing framework (Craik 2002, Craik and Lockhart 1972), which posits that deeper cognitive engagement leads to more durable memory traces (Lindstromberg 2020). The "curiosity account" (Potts and Shanks 2014, Strong 2023) from psychology and applied linguistics further suggests that a lack of information triggers curiosity, motivating learners to seek clues and thereby improving retention even when inferences are imprecise (Kang et al. 2009). When combined with corrective feedback, this process is reinforced through a sense of reward (Potts and Shanks 2014). In discourse, MI also plays a key role in clarifying idioms, a category that often includes PVs due to their figurative and non-transparent nature (Boers 2014, Cooper 1999, Park and Chon 2019). However, its specific effectiveness for PV learning has yet to be established, pointing to an important gap in the literature.

In short, the limited empirical research on how different methods affect both the form and meaning of PVs calls for systematic investigation. This study examines the differential effects of two form-focused methods, particle retrieval (PR) and chunk retrieval (CR), and one meaning-based method, meaning inferencing (MI). Learning outcomes are measured through both form and meaning recall to identify the most impactful method in supporting short- and long-term retention of PV knowledge. Additionally, existing studies have tended to overlook affective variables, such as learners' perceptions and evaluations, which are critical factors in second language learning. Accordingly, the present study also examines learner judgments of the three methods, focusing on preferences, perceived effectiveness, and ease of learning, as gathered through participant surveys and interviews to supplement objective test results. The following research questions therefore guide the present study:

1) Do the three conditions (i.e., PR, CR, MI) differ in their effects on learners' recall of PV meanings in the short and long term?
2) Do the three conditions (i.e., PR, CR, MI) differ in their effects on learners' recall of PV forms in the short and long term?
3) How do learners perceive and evaluate the three conditions (i.e., PR, CR, MI) in terms of preference, perceived effectiveness, and ease of learning?

2. Literature Review

2.1 Form-focused Retrieval Practice for Phrasal Verbs

Explicit vocabulary learning, which deliberately targets vocabulary through retrieval, noticing, and repetition (Laufer 2001, Webb and Nation 2017), has proven effective and warrants further investigation. Unlike implicit learning, which relies on repeated exposure without study intent, explicit learning typically produces faster and greater gains, especially in EFL contexts with limited input and time (Laufer 2005). While traditionally discussed at the level of single words, explicit treatment has also shown benefits for MWUs, whose structural complexity and opaque meanings make form-meaning mapping difficult (Pellicer-Sánchez 2020, Peters 2014). Recent studies have increasingly emphasized targeted form-focused exercises and instructional resources to support the acquisition of MWUs (Boers et al. 2014, Schmitt and Schmitt 2020). Against this background, the present review first considers form-focused retrieval practice as a core instructional approach in L2 vocabulary learning and then turns to how this approach has been applied to phrasal verb learning, with particular attention to particle-focused retrieval.

Barcroft's (2002) Type of Processing Resource Allocation (TOPRA) model suggests that an individual's cognitive resources are limited, such that directing attention toward one language aspect reduces the capacity available for the other. Focusing on form may come at the expense of processing meaning, and vice versa. This cognitive trade-off highlights that different processing types influence distinct aspects of knowledge (Waring and Takaki 2003) and underscores the need to isolate and examine form and meaning separately, even if they are generally intertwined in L2 vocabulary teaching (Laufer 2020).

Retrieval practice is a well-established and highly effective method of explicit L2 vocabulary learning, recognized in both cognitive psychology and language education for its focus on lexical form (Lindstromberg 2020, Roediger and Karpicke 2006). Among retrieval formats, L1-to-L2 cued recall, where learners retrieve an L2 form based on its L1 meaning, has been widely employed in vocabulary research (e.g., Kida and Barcroft 2018, Strong and Boers 2019a). This method is typically implemented through flashcards, gap-filling, or L1-L2 matching exercises, often reinforced by oral repetition, writing, or visualization techniques.

Given their prevalence and significance, PVs and other MWUs have become a central focus in language acquisition research, with growing attention to phrase-level learning rather than isolated single words (Pellicer-Sánchez and Boers 2018). Although PVs are highly salient, making up about one-third of English verbs (Li et al. 2003) and appearing every 150 words on average (Gardner and Davies 2007), they are notoriously difficult to learn due to their non-compositionality and polysemy, with frequent items carrying multiple meanings (e.g., take out = remove, date, or kill). These challenges often lead learners to avoid PVs in favor of simpler one-word verbs (Liao and Fukuya 2004). Such avoidance reduces exposure and results in non-native-like output (Siyanova and Schmitt 2007), thereby reinforcing the cycle of difficulty.

Traditional PV research has predominantly focused on particles (e.g., *up, off*), identifying them as a major source of learner confusion due to their multiple senses and strong influence on verb meaning (Thim 2012, Yasuda 2010). The verb typically provides a stronger semantic clue, rendering particles less salient and more likely to be overlooked, as in blow up or give in. To illustrate the role of retrieval-based practice in PV learning, prior studies have contrasted particle-focused retrieval with particle guessing, a non-inferential procedure in which learners supply particles without prior exposure. For instance, Strong and Boers (2019a) compared a retrieval condition, in which learners first studied definitions and example sentences before recalling missing particles, with a guessing condition, in which learners supplied the missing particle without prior exposure and then received corrective feedback. Japanese university students were randomly assigned to learn 14 target PVs through either retrieval or guessing. In post-tests requiring the missing particle (e.g., hang __ = spend time with friends), the retrieval group consistently outperformed the guessing group. In contrast, the guessing group showed lower retention, as many initial errors resurfaced despite corrective feedback, suggesting that early arbitrary guesses formed persistent incorrect memory traces.

In a follow-up study, Strong and Boers (2019b) tested whether feedback timing influenced learning outcomes. Japanese learners practiced missing-particle tasks under four conditions involving retrieval or guessing, with feedback provided either item by item or after all 14 items. In post-tests requiring learners to produce the full PV (e.g., __ __ = to spend time with friends), a more demanding task, retrieval again outperformed guessing, though the advantage weakened over time due to attrition across groups. Notably, most effective results emerged from retrieval in smaller sets of seven, while feedback timing made little difference.

Taken together, prior research has demonstrated a clear contrast between retrieval and guessing, highlighting the advantage of particle-focused retrieval. Particle guessing can be largely random and inherently relies on arbitrary choices rather than informed reasoning in the absence of prior exposure to target forms. In contrast, retrieval strengthens memory for previously encountered vocabulary by requiring active recall (Nation 2013).

While research on PVs as holistic units remains scarce, such a perspective is well represented in studies on collocations. Insights from collocation research are relevant here because they offer alternative form-focused options within retrieval-based learning. Boers et al. (2014) compared chunk-focused tasks, in which learners selected intact collocations (e.g., *make the bed*), with part-focused tasks requiring reconstruction (e.g., matching or choosing verbs). Among South Asian adult learners, chunk-focused exercises were significantly more effective for form recall, likely due to reduced interference and fewer erroneous associations. In contrast, part-focused formats often led to incorrect combinations that lingered in memory. Building on these findings, Boers et al. (2017) removed two part-focused types and introduced a new format in which learners supplied the verb with only the first letter provided. Results again favored the chunk-focused format, especially for form recall. Ferguson et al. (2024) replicated this line of research with Japanese high school learners and similarly found chunk-focused tasks to be the most beneficial, attributing their effectiveness to reduced cognitive load and enhanced understanding of intact collocation.

These findings point to the value of learning collocations as intact units. Yet, unlike collocations with two relatively transparent elements (e.g., *strong tea*), PVs often rely on opaque particles (e.g., *give in*) that carry essential meaning. Particle-focused learning may therefore be less cognitively demanding than processing full chunks. As most PV studies have centered on particles and few have directly compared chunk- and part-focused methods, further research is needed to clarify whether the benefits observed in collocation learning also apply to PVs.

2.2 Meaning-inferencing Strategy for Multi-word Units

Semantic elaboration involves processing lexical items through their meaning and conceptual content, rather than their surface features, thereby enhancing comprehension (VanPatten et al. 2004). In contrast to the form-focused retrieval approaches reviewed above, meaning-oriented practice can include linking lexical concepts to conceptual frameworks or judging word appropriateness in context (Craik and Tulving 1972). Among such methods, MI—defined as making informed guesses about a word's meaning based on linguistic clues, context, and world knowledge (Haastrup 1991)—is one of the most widely studied strategies over the past several decades. MI is especially valuable given the impossibility of explicitly teaching all vocabulary (Nation 2013), and research indicates that it is the most frequently used and preferred strategy for dealing with unknown words (Schmitt and Schmitt 2020). Although traditionally examined in the context of incidental learning or extensive reading, explicit MI can also function as a deliberate instructional tool when designed to promote reflection on word meanings (Peters 2007).

However, empirical findings remain mixed regarding the effectiveness of this strategy. Some studies report that MI is inefficient and time-consuming, with low success rates and increased risks of confusion for lower-proficiency learners (Calvo et al. 2003, Hamada 2014). For instance, Mondria (2003) found no significant long-term differences between "meaning-inferred" and "meaning-given" methods, although verification phases supported some degree of retention. Conversely, other studies highlight the benefits of MI, including improved retention, learner autonomy, and engagement (Huang and Lin 2014, Yu and Boers 2023). These gains are often explained by the Levels of Processing theory (Craik and Lockhart 1972), which links deeper cognitive engagement to more durable learning outcomes. The curiosity account (Potts and Shanks 2014) further supports MI by proposing that knowledge gaps trigger motivation and drive information-seeking behavior. Yet, the effectiveness of MI depends on timely feedback to prevent the persistence of errors (Elgort et al. 2016), as initial mistakes may form lasting memory traces (Muikku-Werner 2017). Thus, while MI offers cognitive engagement, motivation, and long-term potential, its effectiveness varies depending on learner proficiency, reading experience, and instructional scaffolding. Careful integration with corrective feedback and complementary instructional methods is therefore essential to maximize its pedagogical value.

Most MI studies have focused on single words, and empirical investigations of PV learning remain notably scarce. Some insights can be drawn from idiom studies. For example, Park and Chon (2019) found that Korean learners most frequently used MI and background knowledge to decode unfamiliar idioms, with both strategies showing strong positive associations with idiom recognition scores. Yu and Boers (2023) further demonstrated that inferencing, particularly when supported by information about an idiom's literal origin, led to higher retention than simply providing meanings. These findings underscore both the value of inferencing and the importance of supporting accurate guesses. Given the non-compositional nature of PVs and their reliance on opaque particles, PVs, like idioms, may similarly benefit from MI. However, to the best of our knowledge, no experimental studies have yet systematically applied MI to PV learning. For this reason, the present study incorporated MI alongside two retrieval-based practices (PR and CR) to examine their comparative effects on both form and meaning learning of PVs.

3. Method

3.1 Participants

Thirty-six Korean adults (14 males, 22 females), aged 20 to 51 (M = 31.6, SD = 7.96), participated in the study. Participants were recruited from three English meetup groups and a university in Seoul and were randomly assigned to one of three groups (i.e., PR, CR, MI), each consisting of 12 participants, with a balanced gender distribution to avoid gender-related bias. All participants volunteered to join the study and were recruited through in-person visits to meetup groups and university classes. Those recruited from the university were undergraduate students enrolled in an English Education program, motivated by an interest in improving practical English proficiency, including conversational expressions and high-frequency vocabulary relevant to standardized English tests (e.g., TOEIC, OPIc). The meetup groups were informal, self-directed study groups rather than formal classes, with the primary goal of improving spoken English fluency and everyday vocabulary use. The experiments were scheduled according to individual or group availability and conducted in quiet cafés or university classrooms between March and April 2025.

All participants had received formal English education in Korea, though 41.7% reported beginning intensive English study only after the age of 20. Their primary motivation for learning English was to improve speaking skills for business, daily communication, or test preparation (e.g., TOEIC Speaking, OPIc). Reported official test scores included TOEIC (n = 15, M = 887, SD = 48.2) and OPIc (n = 11; nine at the IH level and two at the AL level).

Before the experiment, participants completed a pretest on the target PVs and the Vocabulary Size Test (VST; Nation and Beglar 2007) to assess general vocabulary proficiency. The average VST score was 68.7 out of 140 (SD = 17.3), equivalent to an estimated vocabulary size of approximately 6,870 word families, indicating upper-intermediate to advanced proficiency. Although the participant age range was relative wide, vocabulary size scores indicated a broadly comparable proficiency level across participants, with no extreme outliers identified. While 61.1% rated vocabulary study as "very important," actual study habits varied: 38.9% reported studying vocabulary "occasionally," 30.6% "rarely," and only a few participants reported engaging in regular weekly vocabulary study.

3.2 Target Phrasal Verbs and Learning Materials

Target PVs were selected primarily based on frequency, a key indicator of pedagogical value in corpus-based research (Liu 2011). High-frequency items were prioritized to reflect authentic usage and provide practical value for learners (Garnier and Schmitt 2015). The study adopted the systematically compiled list of 150 high-frequency PVs developed by Garnier and Schmitt (2015), focusing on the top two senses of each PV, which together account for approximately 75% of actual usage. From this, 44 PVs were selected based on semantic distinctiveness, and six additional items (e.g., *rip off, slack off, call off*) were included from digital sources due to their high salience in spoken discourse, resulting in a 50-item pretest.

Based on the pretest results, 21 target PVs were finalized: *go off, set out, come along, look out, come over, rip off, carry on, wind up, take on, give in, come up, call off, sort out, reach out, turn out, take up, put out, come through, get by, bring about,* and *slack off*. Of these, 17 were sourced from the original list and four from the additional source. Although some overlap in verbs or particles remained—a common characteristic of high-frequency PVs—interference was minimized by assigning similar items (e.g., *take up* vs. *take on*) to different conditions and balancing difficulty across sets. Each condition included seven PVs, consistent with prior studies (Strong 2023) and established limits on cognitive capacity (Miller 1956), as supported by cognitive load theory (Chandler and Sweller 1991). All learning materials were specifically designed based on established MWU exercise procedures reported in previous studies (e.g., Strong and Boers 2019a, Strong and Leeming 2024, Yoo and Yoon 2019). Each learning condition followed a structured, time-controlled sequence of 50 seconds per PV, as determined through pilot testing to ensure cognitive manageability and procedural consistency.

As shown in Figure 1, in the PR condition, participants studied using physical flashcards consisting of six pages per PV. They first viewed the L1-L2 pair (five seconds), read example sentences (15 seconds), recalled and briefly wrote the particle based on the L1 meaning (10 seconds), and then confirmed the correct answer (five seconds). To reinforce the retrieval effect, the recall and confirmation steps were repeated. Seven PVs were studied with a 10-second interval between items, resulting in a total learning time of six minutes. To ensure consistent engagement, flashcards were hand-made and printed on cardstock rather than presented in digital format. Learners worked independently while receiving real-time verbal guidance throughout the procedure. The CR condition followed the same procedure and timing as PR, except that participants recalled and wrote the full PV form rather than only the particle during the recall stages.

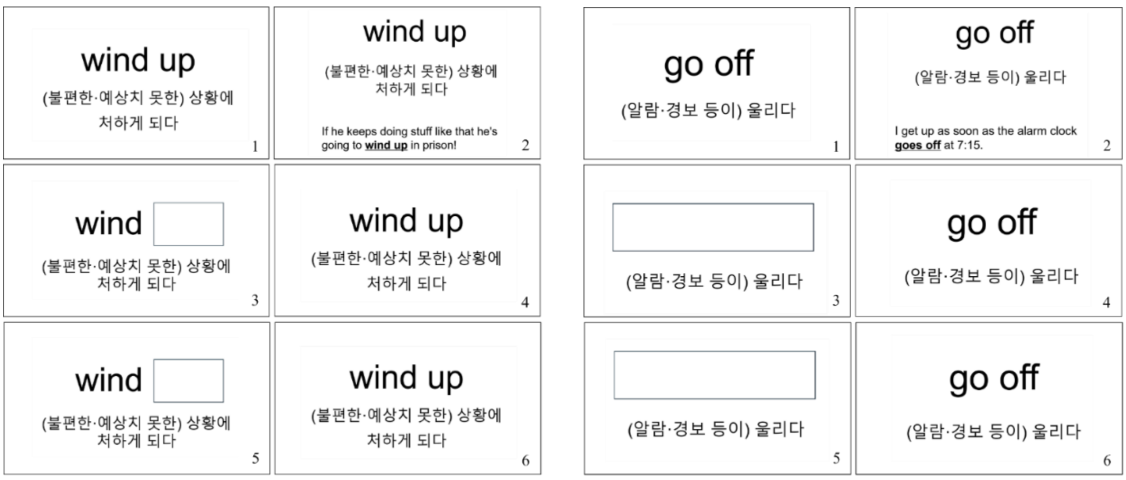

Figure 1. Flashcard Example for PR Practice (left) and CR Practice (right)

For MI, as shown in Figure 2, four-page card sets were used to guide contextual inferencing. Participants briefly viewed the target PV without its meaning (three seconds), then read the first context silently and wrote their inferred meaning in a small box at the bottom right corner (22 seconds). They then repeated the same process with the second context (for another 22 seconds) and were finally given the L1 meaning to verify and confirm the correct answer (for three seconds). Thus, each PV was designed to take a total of 50 seconds to complete, as in PR and CR conditions. To ensure procedural consistency across conditions, participants were instructed to write their inferred meanings in a designated small box on the cards, which was deliberately included to parallel the writing component present in the retrieval tasks.

Figure 2. Flashcard Example for MI

3.3 Pretest and Post-tests

Participants first took the VST (140 items covering the 1K-14K frequency bands) to assess their general vocabulary knowledge, with a 24-minute time limit administered via Google Forms, as vocabulary size can influence inferencing-based learning (Calvo et al. 2003, Hamada 2014). A separate pretest was used to screen prior PV knowledge and finalize the 21 target items. From the initial pool of 50 PVs, items that were unfamiliar, incorrectly answered, or literally translated were retained. The mobile-based test included time tracking, navigation control, and a one-attempt policy to maintain test integrity.

Although each condition targeted either form (PR, CR) or meaning (MI), both immediate and delayed post-tests assessed both form and meaning to avoid methodological bias and to capture potential transfer effects. Immediate posttests (six minutes per condition) followed each learning condition, with form recall tested before meaning recall to prevent semantic priming effects that might artificially enhance performance. For the form recall test, participants were presented with Korean meanings and asked to supply the corresponding PV, with the first letter provided to reduce guessing and ensure accuracy. The test included seven target PVs and three distractors, with all items embedded in new, simplified sentences drawn from major corpus data (e.g., COCA, BNC), and online dictionaries (e.g., Cambridge, Collins). All instructions were given in Korean, and the design followed established practices in L1-to-L2 recall testing (Laufer and Goldstein 2004, Schmitt and Schmitt 2020). For the meaning recall test, participants provided Korean translations for target PVs presented in contextualized sentences. Sentences were sourced from major corpora and refined and standardized for clarity and length. Responses required the present-tense verb form to align with the learning phase. The delayed post-test was administered two weeks later. All three conditions were combined into a single test using the same 21 target items plus six distractors, presented in a new randomized order, with an eight-minute time limit. Participants were instructed not to review target items during the interval.

3.4 Surveys and Interviews

Before the experiment, participants completed a background questionnaire covering demographic information (e.g., age, gender), English learning history (e.g., onset age, study abroad experience), and vocabulary learning habits, which were considered potential factors influencing learning outcomes. After the delayed post-test, a follow-up survey measured learners' perceptions of the three learning conditions using a 4-point Likert scale assessing preference, perceived effectiveness, and ease of learning. Difficulty scores were reverse-coded so that higher values consistently reflected more positive perceptions. Survey items were adapted from prior EFL studies and refined through pilot testing.

To complement the survey data, semi-structured interviews were conducted with six participants selected based on distinctive response patterns and expressed interest in vocabulary learning. Conducted in Korean, each interview lasted under 20 minutes and explored perceived difficulty, usefulness, and individual experiences. These interviews provided qualitative insights into learner attitudes and factors influencing performance.

3.5 Procedure

This study employed a quasi-experimental within-subjects design with 36 adult EFL learners randomly assigned to three groups (n = 12 per group). Each group experienced all three learning conditions (MI, PR, CR) in different sequences to reduce order effects. The sequences were as follows: Group 1: MI → PR → CR; Group 2: PR → MI → CR; Group 3: CR → PR → MI. This arrangement ensured that MI appeared in all three possible positions to control for positional sensitivity, while PR and CR were partially counterbalanced. Each group completed the learning session on a separate day.

Before the main experimental session, participants completed a background survey, consent form, pretest, and VST. During the experiment, participants studied three sets of flashcards (seven PVs per set), each corresponding to a different learning condition. Uniform procedural instructions were provided in Korean by the first author, without explicit instruction or feedback. Each learning condition was followed immediately by form and meaning recall post-tests (two minutes each), with a three-minute break between conditions. Form recall was always administered before meaning recall to prevent contamination from semantic cues. The entire session lasted approximately 50 minutes. Two weeks later, participants completed the delayed post-test assessing long-term retention, followed by a post-experiment survey. All responses were collected for analysis, and selected participants were invited to take part in follow-up interviews to gain further insight into their learning experiences.

3.6 Data Analysis

For research questions 1 and 2, which examined the effects of PR, CR, and MI on form and meaning recall in both immediate and delayed tests, four sets of test responses were scored using predetermined criteria. For form recall, responses received 1 point for correct verb–particle combinations that conveyed the intended meaning, regardless of minor grammatical errors; incorrect particles or meanings received 0 point. For meaning recall, accurate Korean translations or acceptable synonymous responses were awarded 1 point, while all other responses received 0 points. To ensure scoring reliability, a second trained rater independently scored 30% of the data. Inter-rater reliability, assessed using Cohen's Kappa in Jamovi 2.6.26, indicated strong to near-perfect agreement (κ = .82–.97, all ps < .001). Discrepancies were resolved through discussion, after which the first rater scored the remaining data using the agreed-upon criteria.

Descriptive statistics were computed for form and meaning recall at both time points. A 3 (condition) × 2 (time) repeated-measures ANOVA was conducted to examine the main effects of condition and time, as well as their interaction. This analysis allowed for comparison of condition effectiveness, short- and long-term learning effects, and condition-specific retention patterns over time.

For research question 3, which focused on learner perceptions, quantitative data from the post-survey were analyzed to identify trends in preference, perceived effectiveness, and ease of learning. Background survey data provided additional context regarding participants' prior learning experiences and vocabulary strategies, which informed the development of interview questions.

Qualitative data from six semi-structured interviews were transcribed, inductively coded, and analyzed to identify recurring themes. Table 1 summarizes the key codes, representative comments, and the number of participants expressing similar views. Major themes included familiarity with retrieval formats, difficulty with particles, perceived benefits of PR's structure, deeper engagement during MI, cognitive overload under time pressure, and the use of visual and memory strategies. These themes are integrated into the discussion to support the interpretation of the quantitative findings.

Table 1. Coded Themes and Sample Comments from Learner Interviews

Code Theme	Comments	n
Familiarity with PR/CR	"It felt like how I usually studied for tests."	6
Difficulty of particles	"The particle is the trickiest part."	4
Usefulness of PR	"The flashcards layout looked well-organized and clear."	3
	"It helped me focus on and remember the particles."	4
Cognitive load: Confusion of MI	"I felt pressured… It was difficult to guess."	2
Enjoyment: Deeper thinking of MI	"It felt less like memorizing and more enjoyable."	5
Careful reading: Use of contextual clues	"I had to read the example sentences carefully to figure out the meaning."	5
Used personal memory strategies	"I made up my own way to remember the word in my head."	3

4. Results and Discussion

4.1 The Effects of the Three Conditions on PV's Meaning Recall

The first research question examined whether there were any differences in meaning recall scores across the three learning conditions (i.e., PR, CR, MI). Table 2 presents descriptive statistics for PV meaning recall across the three conditions at two time points (immediate and two-week delayed). The skewness and kurtosis values for all conditions fell within acceptable ranges (±2), indicating that the assumption of normality was not violated (Bachman 2004).

As shown in Table 2, all conditions achieved relatively high scores on the immediate post-test, for which the maximum possible score was 7. PR produced the highest mean score (M = 6.28, SD = 0.88), closely followed by MI (M = 6.22, SD = 1.12), while CR showed slightly lower performance (M = 5.94, SD = 1.01). In the delayed post-test, scores declined overall but remained above the midpoint. PR again yielded the highest mean score (M = 5.19, SD = 1.56); however, the ranking shifted, with CR moving to second place (M = 5.08, SD = 1.44), while MI showed the lowest performance (M = 4.81, SD = 1.69). Although these differences were not statistically tested at this stage, the pattern suggests that PR may offer an advantage for promoting long-term retention of PV meanings. The notable increase in SD values from the immediate to the delayed post-test indicates greater variability in learner performance over time. That is, while overall mean scores declined after two weeks, individual differences in retention became more pronounced.

Table 2. Descriptive Statistics of Meaning Recall Scores by Three Conditions

	Condition	M	SD	Min	Max	Skewness	Kurtosis
Immediate	PR	6.28	0.88	4	7	-0.86	-0.43
	CR	5.94	1.01	4	7	-0.58	-0.73
	MI	6.22	1.12	3	7	-1.36	0.92
Delayed	PR	5.19	1.56	2	7	-0.72	-0.18
	CR	5.08	1.44	1	7	-0.76	0.67
	MI	4.81	1.69	1	7	-0.55	-0.37

As shown in Table 3, a two-way repeated measures ANOVA was conducted to examine the main effects of learning condition and time (immediate vs. delayed), as well as their interaction, on participants' meaning recall performance. The assumption of sphericity was met for all effects, as indicated by Mauchly's Test (Condition: W = 0.955, Time: W = 1.000, Condition × Time: W = 0.970, all $p > .05$). The results showed no significant main effect of learning condition on meaning recall scores, $F(2, 70) = 1.16$, $p = .320$, $\eta^2_p = .032$. In other words, no single condition resulted in meaning recall performance that was statistically superior to the others.

Table 3. Repeated Measures ANOVA for Meaning Recall Scores by Three Conditions

	Sum of Squares	df	Mean Square	F	*p*	η^2_p
Condition	2.37	2	1.185	1.16	0.320	0.032
Time	67.78	1	67.782	61.81	< .001	0.638
Condition x Time	2.81	2	1.407	1.95	0.150	0.053

However, a significant main effect of time was found, $F(1, 35) = 67.782$, $p < .001$, $\eta^2_p = .638$, indicating a substantial decline in meaning recall after two weeks, regardless of condition. The interaction between condition and time was not significant, $F(2, 70) = 1.95$, $p = .150$, $\eta^2_p = .053$, suggesting that the degree of decline was similar across all groups.

In summary, all three learning conditions demonstrated statistically similar effectiveness in initial meaning learning, with no significant superiority observed in meaning recall. Importantly, performance declined across all conditions after two weeks, although overall scores remained above the midpoint, suggesting that some degree of retention was maintained. The non-significant main effect of condition and the interaction effect indicate that learners experienced a comparable decline in meaning retention over time.

These findings of similar effectiveness and sustained retention can be partly attributed to the contextualized design of the test items. Since PV meanings are highly context-dependent and MWU learning relies heavily on textual cues, all target vocabulary was embedded in meaningful sentences. To maintain alignment with the learning conditions, contextualized formats were used consistently across tests. As a result, contextual clues may have facilitated meaning recall even when representations were incomplete, contributing to relatively high performance on both the immediate and delayed post-tests.

Interestingly, although not statistically significant, PR outperformed MI and CR in immediate meaning recall (PR > MI > CR), while MI ranked lowest in the delayed test. Despite its semantic focus, MI did not yield a clear recall advantage, possibly due to limited exposure within a single learning session, a factor that the TOPRA model (Barcroft 2002) does not explicitly address. Whereas PR and CR inherently involve repeated retrieval of target forms, MI prioritizes meaning inferencing and subsequent confirmation. Thus, differences in exposure and repetition can be viewed as a natural consequence of each instructional focus, all of which were implemented within equivalent time constraints. These findings suggest that the robust effects of retrieval practice may have outweighed the intended benefits of meaning inferencing, particularly under single-session learning conditions. This interpretation is consistent with previous research demonstrating that meaning-focused tasks alone may not promote long-term retention without deliberate attention to form (Webb and Nation 2017). Participants' written responses during the MI session showed a high initial success rate in inferencing (171 out of 252; 67.86%); however, in the absence of retrieval-based reinforcement, these inferred meanings were less likely to be retained over time.

4.2 The Effects of the Three Conditions on PV's Form Recall

The second research question examined whether there were any differences in the form recall scores across the three learning conditions. Table 4 presents descriptive statistics for PV form recall across the three conditions at two time points. Skewness and kurtosis values fell within acceptable ranges (±2), indicating that the assumption of normality was not violated (Bachman 2004).

Table 4. Descriptive Statistics of Form Recall Scores by Three Conditions

	Condition	M	SD	Min	Max	Skewness	Kurtosis
Immediate	PR	6.17	1.03	3	7	-1.35	1.59
	CR	5.31	1.12	2	7	-0.39	0.83
	MI	4.97	1.56	2	7	-0.38	-0.82
Delayed	PR	2.58	1.52	0	6	0.35	-0.85
	CR	2.50	1.65	0	6	0.26	-0.84
	MI	1.92	1.54	0	6	0.80	0.24

Form recall (max score = 7) showed clearer differences across conditions than meaning recall, corroborating established research findings that form recall is more challenging than meaning recognition or recall and thus a more reliable indicator of learner knowledge (Laufer and Goldstein 2004, Schmitt and Schmitt 2020). PR consistently outperformed CR and MI. In the immediate post-test, PR ($M = 6.17$, $SD = 1.03$) outscored CR ($M = 5.31$, $SD = 1.12$) and MI ($M = 4.97$, $SD = 1.56$). In the delayed post-test, PR ($M = 2.58$, $SD = 1.52$) and CR ($M = 2.50$, $SD = 1.65$) again outperformed MI ($M = 1.92$, $SD = 1.54$). Although all scores declined substantially over time, PR continued to show the highest mean performance, indicating its effectiveness for both immediate and longer-term learning.

To test the statistical significance of these differences, a two-way repeated measures ANOVA was conducted. The assumption of sphericity was met for all effects, as indicated by Mauchly's test (Condition, $W = .955$, $p = .460$; Condition × Time, $W = .970$, $p = .601$). As shown in Table 5, a significant main effect of condition was found, $F(2, 70) = 14.85$, $p < .001$, $\eta^2_p = .298$, indicating that form recall scores differed significantly across conditions with a large effect size. Time also showed a significant main effect, $F(1, 35) = 368.62$, $p < .001$, $\eta^2_p = .913$, confirming a substantial decline in performance across all conditions over two weeks. In addition, a significant condition × time interaction was observed, $F(2, 70) = 3.53$, $p = .035$, $\eta^2_p = .092$, suggesting that the rate of forgetting differed by condition. Specifically, the initial advantage of PR in form recall diminished over time, as it exhibited the largest decline. As a result, the performance gap between PR and CR narrowed after the delay, whereas MI consistently yielded the lowest scores at both time points. As shown in Table 6, Bonferroni-adjusted post hoc comparisons confirmed that PR significantly outperformed both CR ($p = .033$) and MI ($p < .001$), and that CR also outperformed MI ($p = .014$).

Table 5. Two-way Repeated Measures ANOVA for Form Recall Scores by Three Conditions

	Sum of Squares	df	Mean Square	F	p	η^2_p
Condition	31.18	2	15.588	14.85	< .001	.298
Time	535.19	1	535.185	368.62	< .001	.913
Condition x Time	5.68	2	2.838	3.53	.035	.092

Table 6. Post-hoc Comparisons for Form Recall Scores

Comparison	Mean Diff.	SE	t(35)	*p*
PR vs. CR	0.472	0.176	2.69	.033
PR vs. MI	0.931	0.183	5.09	< .001
CR vs. MI	0.458	0.152	3.01	.014

Given the significant Condition × Time interaction effect observed in Table 5, follow-up analyses were conducted to decompose this interaction and examine the pattern of condition differences at each time point. Accordingly, separate one-way repeated measures ANOVAs were performed for the immediate and delayed recall tests to determine whether the learning conditions differed at each time point. For the immediate post-test, Mauchly's Test confirmed that the assumption of sphericity was met (W = .847, p = .059). Table 7 revealed a significant main effect of condition on immediate form recall, $F(2, 70) = 16.70$, $p < .001$, $\eta^2_p = .323$, indicating a very large effect size ($\eta^2_p = .323$). As shown in Table 8, PR significantly outperformed the other two conditions, while no significant difference was observed between CR and MI.

Table 7. One-way Repeated Measures ANOVA for Immediate Form Recall Scores

	Sum of Squares	df	Mean Square	F	*p*	η^2_p
Condition	27.40	2	13.68	16.70	< .001	.323
Residual	57.30	70	0.82			

Table 8. Post-hoc Comparisons for Immediate Form Recall Scores

Comparison	Mean Diff.	SE	t(35)	*p*
PR vs. CR	0.861	0.188	4.59	< .001
PR vs. MI	1.194	0.251	4.75	< .001
CR vs. MI	0.333	0.195	1.71	.290

For the delayed post-test, Mauchly's Test confirmed that the assumption of sphericity was met (W = .987, p = .801). Table 9 shows a significant main effect of condition on form recall two weeks later, $F(2, 70) = 4.59$, $p = .013$, $\eta^2_p = .116$, indicating a moderate effect size. However, as shown in Table 10, Bonferroni post hoc comparisons revealed no significant difference between PR and CR (p = 1.000), suggesting that PR's initial advantage had diminished over time, resulting in comparable long-term outcomes for the two retrieval-based methods.

Table 9. One-way Repeated Measures ANOVA for Delayed Form Recall Scores

	Sum of Squares	df	Mean Square	F	*p*	η^2_p
Condition	9.50	2	4.75	4.59	0.013	.116
Residual	72.50	70	1.04			

Table 10. Post-hoc Comparisons for Delayed Form Recall Scores

Comparison	Mean Diff.	SE	t(35)	*p*
PR vs. CR	0.083	0.253	0.329	1.000
PR vs. MI	0.667	0.232	2.870	0.021
CR vs. MI	0.583	0.234	2.497	0.052

In summary, the results highlight the effectiveness of retrieval-based practice in supporting accurate form recall, with the pattern (PR > CR > MI) reflecting the benefits of directing learners' attention to lexical form. Both retrieval conditions outperformed MI, supporting the TOPRA model's claim that allocating cognitive resources to form enhances the encoding and accessibility of lexical representations, particularly in tasks requiring precise recall (Kida and Barcroft 2018). Specifically, PR's superiority may stem from its focus on particles, which are semantically opaque and often less salient. By isolating particles for targeted practice, PR may increase learners' metalinguistic awareness and support noticing of subtle linguistic cues, thereby facilitating the development of explicit knowledge of PV structure. However, post hoc results showed no significant difference between PR and CR (p = 1.000), indicating that PR's initial advantage was not sustained over time.

CR's relatively weaker performance can be interpreted in light of the Redundancy Effect, a key principle of Cognitive Load Theory (CLT) (Chandler and Sweller 1991). This effect posits that unnecessary or repetitive information can hinder learning because cognitive resources are expended on processing overlapping information alongside essential content. In the CR condition, encountering high-frequency delexical verbs (e.g., *give, take, make*) required learners to process already familiar verb forms in addition to unfamiliar particles or meanings, potentially increasing cognitive load. This overlap may have diverted attention away from the target elements—particularly particles—and contributed to lower form recall accuracy.

4.3 Learners' Perceptions and Evaluations of the Three Learning Conditions

Table 11 presents learners' ratings of preference, perceived effectiveness, and perceived ease for the three learning conditions. In terms of preference (M = 3.22), MI was rated the highest (M = 3.22, SD = 0.87), whereas PR and CR received equally lower ratings (M = 2.72 for both). With respect to perceived effectiveness (max score = 4), MI was also rated the highest (M = 3.08, SD = 0.84), followed by CR (M = 2.83, SD = 0.85) and PR (M = 2.78, SD = 0.76). Regarding perceived ease, PR was considered the easiest (M = 2.75, SD = 0.69), while CR (M = 2.25, SD = 0.69) and MI (M = 2.17, SD = 0.56) were perceived as more difficult. Overall, MI was viewed as the most effective and most preferred learning condition, despite being perceived as the most demanding. PR, in contrast, was rated as the easiest to use, although it was slightly less preferred, while CR consistently ranked in the lower range across all three perceptual measures.

Table 11. Descriptive Statistics of Learners' Perceptions and Evaluations of Conditions

	Condition	M	SD	Min	Max
Preference	PR	2.72	0.91	1	4
	CR	2.72	0.74	1	4
	MI	3.22	0.87	1	4
Effectiveness	PR	2.78	0.76	1	4
	CR	2.83	0.85	1	4
	MI	3.08	0.84	1	4
Perceived Ease[1]	PR	2.75	0.69	2	4
	CR	2.25	0.69	1	4
	MI	2.17	0.56	1	3

Note. [1]1: Not at all, 4: Very much

As shown in Table 12, a repeated measures ANOVA revealed a significant effect of condition on preference, F(2, 70) = 3.89, p = .025, η^2_p = .100, indicating that participants' preferences varied across the three conditions with a moderate effect size. By contrast, no significant effect was found for perceived effectiveness, F(2, 70) = 1.22, p = .302, η^2_p = .034. For perceived ease, however, a significant effect of condition was observed, F(2, 70) = 11.30, p < .001, η^2_p = .244, demonstrating clear differences in how easy participants perceived each condition to be, with a large effect size.

Table 12. Repeated Measures ANOVA for the Learners' Perceptions and Evaluations of Conditions

	Sum of Squares	df	Mean Square	F	*p*	η^2_p
Preference	6.00	2	3.000	3.89	.025	.100
Effectiveness	1.91	2	0.954	1.22	.302	.034
Perceived Ease[1]	7.17	2	3.583	11.30	< .001	.244

Note. [1]1: Not at all, 4: Very much

Although the ANOVA indicated a significant main effect of condition on preference, post-hoc tests revealed no significant pairwise differences after adjustment, as shown in Table 13. MI was rated more favorably than PR and CR; however, the differences were not statistically significant (all $p > .05$).

Table 13. Post-hoc Comparisons for Preference of the Conditions

Comparison	Mean Diff.	SE	t(35)	*p*
PR vs. CR	0.000	0.178	0.00	1.000
PR vs. MI	-0.500	0.231	-2.17	0.111
CR vs. MI	-0.500	0.209	-2.39	0.067

As for perceived ease, post-hoc comparisons reported in Table 14 showed that PR was rated as significantly easier than both CR (t(35) = 4.58, $p < .001$) and MI (t(35) = 4.00, $p < .001$), while no significant difference was observed between CR and MI (t(35) = 0.60, $p = 1.000$).

Table 14. Post-hoc Comparisons for Perceived Ease of the Conditions

Comparison	Mean Diff.	SE	t(35)	*p*
PR vs. CR	0.500	0.109	4.58	< .001
PR vs. MI	0.583	0.146	4.00	< .001
CR vs. MI	0.083	0.140	0.60	1.000

In sum, learners' perceptions and test results showed a contrasting pattern, although the differences were not statistically significant. MI was rated as the most effective but produced the weakest form recall and only moderate meaning recall. PR was the least preferred yet yielded the strongest performance. In the case of CR, low perception ratings were generally aligned with weaker learning outcomes. These findings suggest that learners' perceptions do not necessarily reflect actual performance. More structured, particle-focused practice, such as PR, may feel less engaging but can support stronger recall by reducing extraneous cognitive load.

Interview data revealed that participants were already familiar with retrieval-based practices from school or private classes, often using self-testing strategies such as covering the Korean meaning or the English form to check recall. Although the physical flashcard format was new to some, the method felt intuitive. However, participants 1 and 3 noted that while these strategies were effective for short-term memorization, especially for exams, they did not always transfer to communicative use. Nevertheless, retrieval practice, particularly PR, proved effective for both form and meaning learning of PVs in the present study.

PR's effectiveness may stem from its targeted focus on particles, a particularly challenging component for Korean learners. Prior studies highlight the critical yet opaque role of particles in determining PV meanings (e.g., *give in, catch on*), whereas verbs provide clearer semantic cues and therefore attract more learner attention (Strong and Boers 2019a, Strong and Leeming 2024, Thim 2012, Yasuda 2010). Because particles lack formal salience, learners often overlook them without explicit instructional focus. In line with this observation, four of the six interviewees reported that remembering the correct particle was "the trickiest part of a PV," and stated that PR helped direct their attention to this component. This supports Schmidt's (1990) noticing hypothesis, which posits that conscious attention is essential for intake and long-term retention. By making particles more salient, PR facilitated deeper encoding and proved especially helpful for intermediate to advanced learners who were able to engage in analytic reasoning during recall.

> Having only the particle missing helped me remember the meaning better …. Somehow, I started to mentally categorize the types of particles that follow certain verbs. (Participant 2)

Participant 6 described using memory strategies involving '*take* + particle', reflecting on the particle in *take on a responsibility* to reconstruct meaning. Even when unsure, she generated plausible interpretations, which may have supported conscious encoding. Other participants similarly noted that the clear and structured flashcard layout, consisting of English phrases, Korean translations, and example sentences, functioned like a concise dictionary, aiding memory and fostering a sense of control over their learning. Overall, PR's effectiveness appears to derive from the combination of active retrieval, explicit particle focus, structured design, and manageable cognitive demands.

Notably, component-focused learning operates differently for PVs and collocations. Collocations typically combine two lexically meaningful elements (e.g., *commit suicide, strong tea*), for which intact chunk presentation helps prevent errors arising from incorrect combination. PVs, by contrast, often rely on semantically opaque particles (e.g., *give in*), which impose distinct cognitive demands. Directing attention to particles can therefore reduce processing load and may be more effective than chunk-based approaches commonly recommended for collocations. This distinction lends support to Strong and Boers' (2019a) argument that acquisition strategies should vary across MWU types, given the heterogeneity within the category.

Five of the six interviewees described MI as more enjoyable, meaningful, and useful, stating that it made them "think more deeply," "use their brain," and experience "aha moments" when inferring or verifying meanings. This engagement enhanced their subjective learning experience, although many acknowledged that they focused more on understanding example sentences than on encoding the PVs themselves. Because all learning conditions were time-controlled, MI provided fewer opportunities for explicit retrieval than PR or CR, which may account for its weaker recall outcomes despite fostering richer contextual understanding.

This trade-off aligns with the "testing effect," which demonstrates that active retrieval plays a more decisive role in long-term retention than contextual processing alone (Webb and Nation 2017). While MI promoted semantic problem-solving and deeper engagement, it diverted attentional resources away from encoding verb-particle combinations as precise lexical forms. As Participant 5 explained, MI led him to prioritize meaning, whereas PR and CR emphasized specific forms. This pattern is consistent with the TOPRA model (Barcroft 2002), which highlights the limits of cognitive resources: effort invested in inferencing may weaken precise form encoding in the absence of explicit form–meaning mapping.

Moreover, MI appeared to be more sensitive to individual differences than the other conditions. Learners with stronger literacy habits, such as participating in book clubs, reading articles, or searching for information online, both performed better and expressed greater enjoyment during inferencing process, showing high accuracy in their flashcard entries. Survey responses further indicated that learners who preferred MI often reported reading-based vocabulary learning experiences. Participants 4 and 6, who regularly used monolingual dictionaries or delayed dictionary consultation to prioritize context guessing, similarly reported both preference for and success with MI. For these learners, successfully inferring meaning boosted confidence and memorability, even when initial guesses were inaccurate.

By contrast, learners with weaker reading habits or lower confidence found MI stressful under time constraints. Participant 1, who had previously studied PVs for TOEIC Speaking but reported reduced motivation after entering the workforce, answered only two MI items correctly and openly discussed her difficulties.

> I would say it (MI) was very difficult for me ... I felt pressured, which made it difficult to figure out the meaning of the sentences … I was especially challenged by the time limit during the session and had a hard time processing the inferencing quickly. (Participant 1)

She explained that limited time and incorrect guesses induced anxiety, leaving her uncertain whether such experiences facilitated or hindered learning. She reported that having clear Korean definitions and example sentences, as provided in PR and CR, felt more supportive by clarifying meanings and usage patterns. These findings suggest that the "curiosity account" (Murayama et al. 2019, Potts and Shanks 2014, Strong 2023)—the idea that knowledge gaps stimulate motivated inference—may not apply uniformly across learners. For participant 1, imprecise guesses, limited prior knowledge, and discomfort with ambiguity reduced effective engagement. Overall, while MI can be cognitively stimulating, its effectiveness as a primary learning strategy appears to depend on individual characteristics such as literacy background, learning preferences, and tolerance for ambiguity.

5. Conclusion

This study investigated the effects of three learning approaches (PR, CR, MI) on the acquisition of English PVs. The findings can be summarized as follows. First, all three methods produced comparable outcomes for meaning recall, with no condition showing significant long-term advantages. Second, PR consistently supported the most accurate form recall, although its superiority diminished over time as CR approached a similar level of performance. Third, learners' perceptions did not align with actual performance. MI was most favored but least effective in recall, while PR was least favored but yielded the strongest outcomes. These results underscore the value of retrieval practice, particularly particle-focused retrieval, for strengthening the structural dimensions of PV learning, while also indicating that learner traits (e.g., reading experience) moderate the effectiveness of meaning-inferencing strategies. PR proved most effective for form recall despite low preference, suggesting that effortful retrieval of difficult elements such as particles promotes deeper encoding. CR, while less effective initially, resulted in comparable long-term outcomes and aligns with real-life communicative needs by reinforcing PVs as holistic units.

In contrast, MI generated strong learner preference and engagement but did not lead to superior retention, indicating the need for additional scaffolding, repeated exposure, or integration with retrieval-based practice. Overall, effective PV pedagogy should balance structured retrieval with contextual richness, such as using PR to secure accurate form-meaning mappings, CR to support fluency in authentic contexts, and MI to foster motivation and deeper engagement. Such a blended approach may help learners overcome the inherent challenges of PVs and promote durable, communicative mastery.

While this study provides valuable insights into the effects of three learning approaches on PV acquisition, several limitations should be acknowledged. First, the relatively small sample size limited statistical power and generalizability. Future studies should include larger samples to strengthen external validity. Second, both form and meaning recall tests embedded target PVs in contextual sentences, which may have inflated meaning recall scores through contextual guessing. Although contextualization was necessary to reflect authentic PV usage, future research could compare contextualized and decontextualized test formats to better isolate recall effects. Third, while learning conditions were counterbalanced across groups, vocabulary items were not rotated between conditions. Despite careful matching of item sets by transparency and difficulty, future work should implement full item-level counterbalancing to more rigorously isolate treatment effects. Nevertheless, this study advances PV research by integrating structural and semantic approaches and by offering a more comprehensive perspective than prior work. Further research with larger samples, refined assessment formats, and stronger methodological controls is needed to extend and validate these findings.

References

Bachman, L. F. 2004. *Statistical Analyses for Language Assessment Book*. Cambridge: Cambridge University Press.

Barcroft, J. 2002. Semantic and structural elaboration in L2 lexical acquisition. *Language Learning* 52(2), 323-363.

Bjork, R. A. and J. F. Kroll. 2015. Desirable difficulties in vocabulary learning. *The American Journal of Psychology* 128(2), 241-252.

Boers, F. 2014. Idioms and phraseology. In J. Littlemore and J. Taylor, eds., *The Bloomsbury Companion to Cognitive Linguistics*, 185-202. London: Bloomsbury.

Boers, F., T. C. T. Dang and B. Strong. 2017. Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research* 21(3), 362-380.

Boers, F., M. Demecheleer, A. Coxhead and S. Webb. 2014. Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research* 18(1), 54-74.

Boers, F., J. Eyckmans and H. Stengers. 2006. Motivating multiword units: Rationale, mnemonic benefits, and cognitive style variables. *EUROSLA Yearbook* 6(1), 169-190.

Calvo, M. G., A. Estevez and M. G. Dowens. 2003. Time course of elaborative inferences in reading as a function of prior vocabulary knowledge. *Learning and Instruction* 13(6), 611-631.

Candry, S., I. Elgort, J. Deconinck and J. Eyckmans. 2017. Word writing vs. meaning inferencing in contextualized L2 vocabulary learning: Assessing the effect of different vocabulary learning strategies. *Canadian Modern Language Review* 73(3), 293-318.

Chandler, P. and J. Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction* 8(4), 293-332.

Cooper, T. C. 1999. Processing of idioms by L2 learners of English. *TESOL Quarterly 33*(2), 233-262.

Craik, F. I. 2002. Levels of processing: Past, present... and future? *Memory* 10(5-6), 305-318.

Craik, F. I. and R. S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior* 11(6), 671-684.

Craik, F. I. and E. Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104(3), 268-294.

Elgort, I., S. Candry, T. J. Boutorwick, J. Eyckman and M. Brysbaert. 2016. Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics* 39(5), 646-667.

Ellis, N. C. 2012. Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics* 32, 17-44.

Ellis, N. C., R. I. T. A. Simpson-Vlach and C. Maynard. 2008. Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly* 42(3), 375-396.

Ferguson, P., A. Siyanova-Chanturia and P. Leeming. 2024. Impact of exercise format and repetition on learning verb-noun collocations. *Language Teaching Research* 28(5), 1750-1776.

Gardner, D. and M. Davies. 2007. Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly* 41(2), 339-359.

Garnier, M. and N. Schmitt. 2015. The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research* 19(6), 645-666.

Garnier, M. and N. Schmitt. 2016. Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System* 59, 29-44.

Haastrup, K. 1991. *Lexical Inferencing Procedures or Talking about Words: Receptive Procedures in Foreign*

Language Learning with Special Reference to English. Tübingen, Germany: Gunter Narr.

Hamada, M. 2014. The role of morphological and contextual information in L2 lexical inference. *The Modern Language Journal* 98(4), 992-1005.

Huang, L. L. and C. C. Lin. 2014. Three approaches to glossing and their effects on vocabulary learning. *System* 44, 127-136.

Jiang, N. 2002. Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 24(4), 617-637.

Kang, M. J., M. Hsu, I. M. Krajbich, G. Loewenstein, S. M. McClure, J. T. Wang and C. F. Camerer. 2009. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science* 20(8), 963-973.

Kida, S. and J. Barcroft. 2018. Semantic and structural tasks for the mapping component of L2 vocabulary learning: Testing the TOPRA model from a new angle. *Studies in Second Language Acquisition* 40(3), 477-502.

Laufer, B. 2001. Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect* 16, 44-54.

Laufer, B. 2005. Focus on form in second language vocabulary learning. In S. H. Foster-Cohen, M. Garcia-Mayo and J. Cenoz, eds., *EUROSLA Yearbook,* Vol. 5, 223-250. Amsterdam: John Benjamins.

Laufer, B. 2020. Evaluating exercises for learning vocabulary. In S. Webb, ed., *The Routledge Handbook of Vocabulary Studies,* 351-360. London: Routledge.

Laufer, B. and Z. Goldstein. 2004. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning* 54(3), 399-436.

Lewis, M. 1993. *The Lexical Approach*. Hove, UK: Language Teaching Publications.

Li, W., X. Zhang, C. Niu, Y. Jiang and R. K. Srihari. 2003. An expert lexicon approach to identifying English phrasal verbs. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics,* 513-520. Sapporo, Japan.

Liao, Y. and Y. J. Fukuya. 2004. Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning* 54(2), 193-226.

Lindstromberg, S. 2020. Intentional L2 vocabulary learning. In S. Webb, ed., *The Routledge Handbook of Vocabulary Studies*, 240-254. London: Routledge.

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

Miller, G. A. 1956. Information and memory. *Scientific American* 195(2), 42-47.

Mondria, J. A. 2003. The effects of inferring, verifying, and memorizing on the retention of L2 word meanings: An experimental comparison of the "meaning meaning-inferred method" and the "meaning-given method." *Studies in Second Language Acquisition* 25(4), 473-499.

Muikku-Werner, P. 2017. Lexical inferencing and the mutual intelligibility of Estonian and Finnish. *Nordic Journal of Linguistics* 40(2), 201-223.

Murayama, K., L. FitzGibbon and M. Sakaki. 2019. Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review* 31(4), 875-895.

Nation, I. S. P. 2013. *Learning Vocabulary in Another Language* (2nd ed). Cambridge: Cambridge University Press.

Nation, I. S. P. 2024. Re-thinking the principles of (vocabulary) learning and their applications. *Languages* 9(5), 160.

Nation, I. S. P. and D. Beglar. 2007. A vocabulary size test. *The Language Teacher* 31, 9-13.

Park, J. and Y. V. Chon. 2019. EFL learners' knowledge of high-frequency words in the comprehension of idioms: A boost or a burden? *RELC Journal* 50(2), 219-234.

Pellicer-Sánchez, A. 2020. Learning single words vs. multiword items. In S. Webb, ed., *The Routledge Handbook of Vocabulary Studies*, 158-173. London: Routledge.
Pellicer-Sánchez, A. and F. Boers. 2018. Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia and A. Pellicer-Sánchez, eds., *Understanding Formulaic Language: A Second Language Acquisition Perspective*, 153-173. London: Routledge.
Peters, E. 2007. Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning and Technology* 11(2), 36-58.
Peters, E. 2014. The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research* 18(1), 75-94.
Potts, R. and D. R. Shanks. 2014. The benefit of generating errors during learning. *Journal of Experimental Psychology: General* 143(2), 644-667.
Roediger III, H. L. and J. D. Karpicke. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science* 17(3), 249-255.
Schmidt, R. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11, 129-158.
Schmitt, N. and D. Schmitt. 2020. *Vocabulary in Teaching*. Cambridge: Cambridge University Press.
Siyanova-Chanturia, A. and N. Schmitt. 2007. Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics* 45, 119-139.
Strong, B. 2023. The impact of guessing and retrieval strategies for learning phrasal verbs. *International Review of Applied Linguistics in Language Teaching* 47(3), 391-420.
Strong, B. and F. Boers. 2019a. The error in trial and error: Exercises on phrasal verbs. *TESOL Quarterly* 53(2), 289-319.
Strong, B. and F. Boers. 2019b. Weighing up exercises on phrasal verbs: Retrieval versus trial-and-error practices. *The Modern Language Journal* 103(3), 562-579.
Strong, B. and P. Leeming. 2024. Comparing the effectiveness of verb-focused and particle-focused exercise formats on the recall and recognition of phrasal verbs. *Language Teaching Research* 28(6), 1801-1824.
Thim, S. 2012. *Phrasal Verbs: The English Verb-Particle Construction and Its History* (Vol. 78). Berlin: De Gruyter.
VanPatten, B., J. Williams, S. Rott and M. Overstreet, eds. 2004. *Form-meaning Connections in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.
Waring, R. and M. Takaki. 2003. At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language* 15(2), 130-163.
Webb, S. and P. Nation. 2017. *How Vocabulary is Learned*. Oxford: Oxford University Press.
Wray, A. 2002. *Formulaic Language and the Lexicon*. 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.
Yoo, S. and H. Yoon. 2019. Effects of word-writing in contextual word learning on Korean high school students' L2 vocabulary learning. *Korean Journal of English Language and Linguistics* 19(4), 713-735.
Yu, X. and F. Boers. 2023. Inferring the meaning of idioms: Does accuracy matter for retention in memory? *RELC Journal* 55(3), 721-734.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

