WWW.PEGEGOG.NET RESEARCH ARTICLE

The Role of Learner Reflection and Error Recollection in Developing Phrasal Verb Knowledge

Dr. Mohammed Hassan Alshaikhi, University of Tabuk

Abstract
This study investigates how two types of gap-fill exercises; copying and trial and error affect the acquisition of phrasal verbs among intermediate Saudi EFL learners. Going beyond traditional task-based learning approaches, the research introduces a novel integration of cognitive reflection and memory retrieval theory into the design of language exercises. Specifically, it examines whether learners' ability to recall their initial guesses during trial-and-error practice influences later performance on both receptive and productive posttests. A mixed-methods design was employed, with participants completing gap-fill tasks followed by immediate and delayed posttests in cued-recall and multiple-choice formats. The results indicate that while copying leads to higher immediate performance, trial and error exercises supported by reflective retrieval promote deeper, longer-lasting learning. Moreover, the ability to remember initial guesses significantly improved the likelihood of retrieving the correct phrasal verb later. These findings offer practical implications for curriculum design, suggesting that allowing learners to engage with errors and reflect on their responses can enhance vocabulary acquisition. The study contributes to second language acquisition research by linking memory-based frameworks with pedagogical tasks, particularly in the underexplored area of phrasal verb learning.
Keywords: phrasal verbs, trial-and-error learning, reflection, Saudi EFL learners, mixed-effects modeling, metacognition, vocabulary instruction, episodic memory

1. Introduction

Phrasal verbs are widely recognized as a frequent and integral part of the English language (Gardner and Davies 2007). While they appear across all types of discourse, they are especially common in informal spoken and written contexts (Liu 2011; Liu and Myers 2020). As a result, learners of English as a second language are likely to encounter phrasal verbs often when engaging with such informal registers (Cornell 1985; Darwin and Gray 1999). However, these expressions are known to present considerable difficulty for language learners (Condon 2008; Garnier and Schmitt 2015; Side 1990), raising questions about whether learners fully comprehend the phrasal verbs they encounter and whether they recognize their importance in producing natural and fluent English.

To address these issues, previous studies have investigated learner understanding of phrasal verbs by asking them to complete sentences containing gaps and to decide whether a phrasal verb or a single word alternative was more appropriate (Dagut and Laufer 1985; Hulstijn and Marchena 1989; Laufer and Eliasson 1993; Liao and Fukuya 2004). Findings from this line of research consistently show that second language learners tend to prefer single word verbs, while native speakers opt more often for phrasal verbs. This pattern suggests that many learners do not have a solid grasp of phrasal verb usage, and their avoidance may lead to language that does not reflect the natural speech patterns of native users (Siyanova and Schmitt 2007).

Given these observations, it is not surprising that numerous scholars have emphasized the need for research focused on how to improve learners' command of phrasal verbs (AlOtaibi 2019; Condon 2008; Sonbul El Dakhs and AlOtaibi 2020; Strong 2013; Strong and Boers 2019a 2019b). The current study responds to these calls by examining alternative ways of implementing gap fill exercises, which are commonly found in English as a second language teaching materials. Focusing on university level learners in Saudi Arabia, this research investigates how different versions of this task can support the development of both receptive and productive knowledge of phrasal verbs in the local EFL context.

Focusing on university-level learners in Saudi Arabia, this research investigates how different versions of this task can support the development of both receptive and productive knowledge of phrasal verbs in the local EFL context. The present study builds upon the design and findings of Strong and Boers (2019a), adapting their trial-and-error vs. copying exercise format to a Saudi university EFL context.

2. Literature Review

2.1. Challenges in Learning Phrasal Verbs

Phrasal verbs are widely acknowledged as one of the most difficult types of multiword expressions for language learners to master (Yasuda 2010). This difficulty stems from several interrelated factors. First, the meaning of a phrasal verb often cannot be easily inferred even when the individual words are familiar. For example, knowing the meanings of run and up does little to help learners understand the phrase run up in the context of run up a bill. In cases like this, the expression functions much like an idiom, requiring more than just vocabulary knowledge to interpret (Kovecses and Szabo 1996).

A second challenge is that many phrasal verbs are highly polysemous. As Gardner and Davies (2007) noted, the most frequently used phrasal verbs in English typically carry an average of more than five distinct meanings. This presents a considerable obstacle for learners, as understanding one meaning of a phrasal verb does not necessarily provide insight into its other uses. Consider, for instance, the verb *make up*, which can appear in diverse contexts such as *make up a story, make up after an argument, make up one's face, make up the difference, make up a bed, make up for something, and make it up to someone*. The connection between these meanings is not immediately obvious, even to advanced learners.

The third major area of difficulty involves the flexibility of word order in phrasal verbs. Although a phrasal verb consists of a verb and a particle that work together as a single semantic unit, the particle is often movable. For example, it is grammatically correct to say both turn off the alarm and turn the alarm off when the object is relatively short. However, inserting longer objects into this structure depends on several factors, including context, rhythm, and the intended meaning (Schmitt and Redwood 2011). Additionally, learners must be aware that when the object is a pronoun, particle separation becomes obligatory, as in turn it off versus the incorrect form *turn off it.

Given these various challenges, language educators and textbook developers have increasingly recognized the need to focus on phrasal verb instruction. As a result, many English language learning materials now include targeted exercises to support the acquisition of these complex yet essential elements of English vocabulary.

2.2. Learning Phrasal Verbs Through Trial and Retrieval Based Tasks

Strong and Boers (2019a) conducted a comprehensive review of how phrasal verbs are treated in 44 widely used ESL textbooks. They discovered several formats, such as gap-fills with missing verbs or particles. These exercises were presented in different instructional contexts. In some cases, learners were encouraged to attempt the exercises

first and then check their answers using a key. In other cases, students were expected to review the target phrasal verbs before completing the task. Alternatively, learners might be prompted to consult earlier parts of the textbook for help during the activity itself.

Building on these observations, Strong and Boers designed an empirical study comparing two of these instructional approaches—retrieval practice and trial and error—with a sample of 140 Japanese EFL learners. In the trial-and-error condition, students were presented with short dialogues in which they had to fill in missing particles for phrasal verbs. If they were unsure of the correct answer, they were encouraged to make a guess. After completing the task, they received feedback in the form of full phrasal verbs embedded in dialogues similar to those in the original activity. In the retrieval condition, learners were first asked to study the same feedback provided to the trial-and-error group, and then complete the same gap fill task.

The impact of both methods was assessed through a delayed cued recall posttest, in which participants were asked to recall the missing particles of the phrasal verbs. The findings revealed that learners in the retrieval group outperformed those in the trial-and-error group, indicating that studying the correct forms in advance supported stronger long-term retention. However, the researchers also analyzed the nature of the errors produced in the trial-and-error condition and found that around 25 percent of the incorrect responses made during the initial exercise phase were repeated in the posttest. This suggests that generating incorrect forms can lead to proactive interference, which in turn may hinder accurate recall and learning.

These findings raise important questions about the effectiveness of error-based learning activities and highlight the need to better understand how feedback and task design influence vocabulary acquisition, particularly for multiword items such as phrasal verbs (Strong and Boers, 2019a).

2.3. Further Insights into Feedback Timing and Error Interference

Although the earlier study by Strong and Boers (2019a) provided useful insights into how different exercise types affect phrasal verb learning, it did not fully explore the role of feedback timing in the learning process. To address this gap, Strong and Boers (2019b) conducted a follow-up study with a new cohort of Japanese learners of English, aiming to examine how the timing of feedback influences learning outcomes across both retrieval and trial and error conditions.

In the retrieval-based conditions, participants either studied one phrasal verb at a time followed by an individual gap fill activity for that item, or they reviewed a group of 14 phrasal verbs in sequence and then completed an exercise covering the full set. Similarly, in the trial-and-error conditions, one group completed a single item gap fill followed by immediate feedback for each item, while another group worked through a full set of 14 gap fill items before receiving cumulative feedback.

Results from a delayed cued recall posttest, where participants were asked to recall the full phrasal verb, revealed that learners in the retrieval conditions performed better overall, regardless of whether the feedback was provided immediately after each item or after the entire set. These results reinforce the conclusion that attempting to guess without adequate support can lead to lasting errors, and that retrieval-based learning offers more consistent benefits for retention.

Importantly, the study highlighted the persistent challenge of error interference. Even with structured feedback, incorrect initial guesses in trial-and-error conditions were shown to negatively affect later recall. This underscores the importance of designing vocabulary exercises that minimize the harmful impact of incorrect responses and provide learners with opportunities for accurate form retrieval and reinforcement. More targeted research is still needed to determine how instructional design can best support this goal, particularly for complex lexical items like phrasal verbs.

2.4. The Impact of Initial Incorrect Guesses on Vocabulary Learning

In recent years, the role of guessing in language learning has gained significant attention, particularly within the field of memory research. A growing body of evidence suggests that making an initial error followed by corrective feedback can actually enhance learning, especially when the items being learned are semantically related, such as in the word pair tide and beach. However, this benefit does not appear to extend to unrelated pairs like stem and candy (Grimaldi and Karpicke 2012; Hays, Kornell, and Bjork 2013; Kornell, Hays, and Bjork 2009). When a learner encounters a cue and guesses incorrectly, receiving the correct answer as feedback may form a mental chain. Later, when the same cue is presented again, it may trigger the memory of the error, which in turn activates the correct answer.

While semantic similarity may explain some learning benefits, it does not account for findings where incorrect guesses enhance vocabulary learning even in the absence of shared meaning. For instance, learners have shown improved recall for word pairs involving a foreign word and its English translation, in cases where no clear semantic link exists. This suggests that additional cognitive mechanisms may support learning from errors, particularly in second language contexts where associations often rely on episodic rather than semantic connections.

While these findings emphasize the potential value of errors in forming memory links, they do not fully explain the role of *conscious recollection*. The next section addresses this issue by focusing on how the act of remembering earlier guesses, rather than just making them, shapes vocabulary acquisition.

2.5. Episodic Memory and the Role of Recall in Error-Based Learning

Extending beyond semantic associations, recent research has shown that the benefits of incorrect guessing are also linked to learners' ability to recall and reflect on those errors. This episodic recollection mechanism differs from earlier models of semantic mediation and highlights the role of memory trace reactivation in learning.

Studies have found that making and correcting errors can also enhance the learning of word pairs across different languages, including first and second language combinations (Potts, Davies, and Shanks 2019; Potts and Shanks 2014; Seabrooke, Hollins, Kent, Wills, and Mitchell 2019). In particular, Metcalfe and Huelser (2020) explored how episodic memory contributes to learning from errors. Their study demonstrated that individuals with impaired episodic memory struggled to benefit from incorrect guesses, as they were unable to recall their previous mistakes. This implies that the ability to mentally retrieve an earlier error plays a key role in successfully updating and encoding the correct response.

In their experiment, participants were asked to study word triplets such as wrist, palm, and hand. They were first presented with two words and prompted to guess the missing third word. After seeing the correct answer, they later completed a recall test that required them to produce the target word again, while also recalling their original guess. The results showed that participants who were able to remember their initial incorrect guesses were more successful at retrieving the correct answers. This suggests that episodic memory, not just semantic links, plays a key role in learning from errors.

2.6. The Impact of Copying Exercises on Vocabulary Learning

As discussed earlier, phrasal verb activities are presented in various formats across ESL instructional materials. In more specialized resources such as *Phrasal Verbs in Use* by McCarthy and O'Dell (2017), the instructional design often places explanations of target expressions on one page and related exercises directly across from them. This layout makes it easy for learners to complete the exercises by simply copying the phrasal verbs they see on the adjacent page.

Nonetheless, broader research into vocabulary learning has examined how writing or copying words influences acquisition. A number of studies have shown that writing words can be more beneficial than passive studying when it comes to learning new or foreign

vocabulary (Candry, Elgort, Deconinck, and Eyckmans 2017; Thomas and Dieter 1987). For instance, Webb and Piasecki (2018) conducted a study with 38 English language learners, who were asked to learn pairs of pseudowords and images under three different conditions: no writing, timed writing, and untimed writing.

In the no writing condition, learners viewed the pseudoword-image pairs and studied them for six seconds each. In the writing condition, learners were required to write the pseudowords during the same time span. A third group was allowed to write the words without any time restriction. The findings revealed that participants who had unlimited time to write and review the words outperformed the others on posttests. This suggests that copying can aid vocabulary learning, particularly when learners have sufficient time to engage actively with the form and meaning of the items.

2.7. The Present Study

As discussed in the previous sections, there has been growing interest in how different formats of gap fill exercises can support the learning of phrasal verbs. However, several important areas remain underexplored. First, despite the common use of gap fill tasks in language classrooms, no prior study has evaluated their effectiveness in developing learners' receptive knowledge of phrasal verbs. Previous research, including the work of Strong and Boers (2019a, 2019b), has focused primarily on productive outcomes, using cued recall tests to assess learners' ability to produce target expressions. To fill this gap, the current study incorporates both a cued recall test and a multiple-choice test to capture gains in both productive and receptive knowledge.

Second, while copying answers directly from instructional materials is a widespread classroom practice, especially in ESL and EFL textbooks, its impact on learning phrasal verbs has not been empirically tested. Although copying is often assumed to support memorization, concrete evidence is needed to validate this approach in the context of phrasal verb acquisition.

Third, there has been limited exploration of how to enhance learning within trial-and-error exercises. One notable exception is the study by Strong and Boers (2019b), which examined the timing of feedback delivery. A promising area for further investigation is the effect of asking learners to recall their initial incorrect guesses. Metcalfe and Huelser (2020) found that learners who were able to remember their earlier incorrect responses were more likely to retrieve the correct answers later on. However, this has not yet been tested in the context of phrasal verb learning.

In response to these gaps, the present study focuses on Saudi EFL learners and aims to answer the following research questions:

1. Which format of a gap fill exercise (trial and error versus copying) is more effective in promoting productive knowledge of phrasal verbs at immediate and delayed testing intervals?
2. Which format of a gap fill exercise (trial and error versus copying) is more effective in promoting receptive knowledge of phrasal verbs at immediate and delayed testing intervals?
3. Does prompting learners in the trial-and-error condition to recall their initial guesses during posttests influence their ability to retrieve the correct phrasal verbs?

To consolidate these insights, the present study draws on a conceptual framework that integrates trial-and-error practice, learner reflection, and phrasal verb acquisition in EFL settings. The acquisition of phrasal verbs through trial and error can be understood through key principles in cognitive psychology, particularly those related to retrieval effort and memory formation. When learners engage in trial-and-error tasks, they attempt to retrieve answers from memory, even if the responses are incorrect. This act of retrieval leaves a mental trace of the learning episode. When learners later reflect on this attempt, especially through follow-up activities like cued recall or guided correction, the memory becomes

more deeply processed and better retained. This process aligns with the Generation Effect, which shows that learners remember information more effectively when they generate it themselves rather than simply receive it from a source. It also draws on the Episodic Encoding Framework, which highlights the role of recalling one's own earlier attempts as a foundation for meaningful learning. In the case of English phrasal verbs, which are often difficult to interpret because their meaning cannot be inferred from individual words, this process helps learners gradually notice, experiment with, and internalize the combinations of verbs and particles in a lasting and meaningful way.

3. Conceptual Framework

Research on phrasal verb learning has increasingly drawn from cognitive theories that explain how learners process, store, and retrieve vocabulary. Two frameworks are central to the present study: the Generation Effect and Episodic Memory Theory. Together, they provide a theoretical basis for understanding how task design influences long-term vocabulary acquisition.

The Generation Effect refers to the enhanced retention that occurs when learners actively generate responses rather than receive them passively (Slamecka & Graf, 1978). When learners are asked to guess or retrieve a word, even if their response is incorrect, the act of generating it leaves a stronger cognitive imprint than simply copying a model. This process is especially relevant in trial-and-error learning, where learners attempt to supply a missing form and later compare it with the correct version.

Episodic Memory Theory, as advanced by Metcalfe and Huelser (2020), adds an important dimension. Their work shows that recalling a prior mistake strengthens the encoding of the correct response. In language learning tasks, when learners reflect on their earlier incorrect guesses, they create meaningful links between the error and the corrected form. This retrieval-reflection loop contributes to deeper learning.

In the context of phrasal verb acquisition, these frameworks support the idea that trial and error learning, when paired with opportunities for error recollection, may enhance both form-meaning mapping and long-term retention. By contrast, copying tasks may promote short-term familiarity but involve less cognitive elaboration, leading to weaker long-term effects.

This conceptual model also incorporates the role of metacognitive reflection. Learners who are asked to recall their guesses engage in metacognitive monitoring, which has been linked to increased awareness and retention of new vocabulary. Thus, the model views vocabulary acquisition as a process involving initial retrieval, feedback, memory encoding, and reflective recall.

Figure 1. Conceptual Model of Trial and Error Driven Phrasal Verb Acquisition

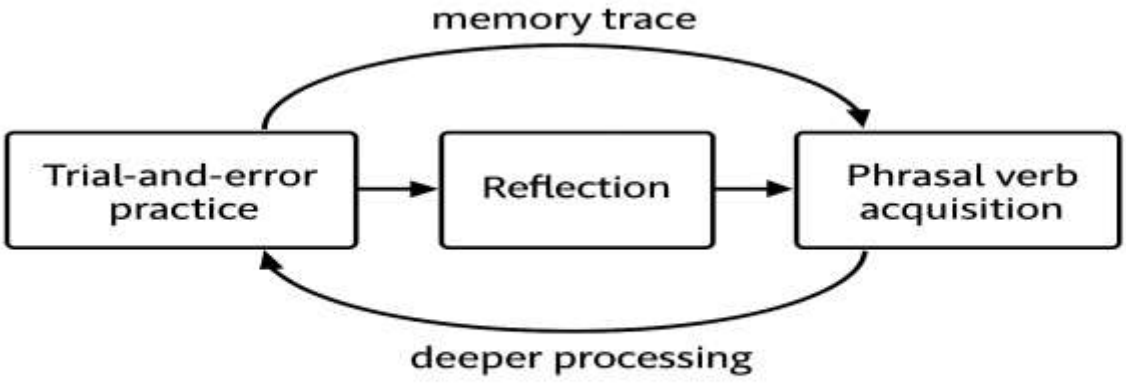

FIGURE 1. Conceptual model of trial and error driven phrasal verb acquisition

4. Methods

4.1. Participants

The participants in this study were 135 Saudi undergraduate EFL learners enrolled in a general English language course at two public universities in Saudi Arabia. The sample included 82 male and 53 female students, aged between 18 and 21 (M = 19.3, SD = 0.71). All participants reported having studied English formally for at least six years (M = 7.8, SD = 2.5) and had recently completed the Oxford Placement Test (OPT), with an average score of 62.3 (SD = 10.4), suggesting an intermediate level of proficiency (B1–B2 range on the CEFR scale).

Participants were drawn from seven intact classrooms, with instructional interventions embedded into their regular coursework. The phrasal verb learning tasks aligned with the syllabus for their vocabulary development module. Prior to participation, students completed a consent form confirming their voluntary involvement and were informed that withdrawal at any stage would not affect their academic standing or grades.

4.2. Materials

The target learning items comprised 24 English phrasal verbs**,** selected from the PHaVE List (Garnier & Schmitt, 2015), which ranks the most commonly used phrasal verbs in English. Several criteria informed the selection process. First, the total number of phrasal verbs was adjusted to fit the 50-minute time frame allocated for treatment and posttests within a standard university class session. Second, the instructional format mirrored that found in many commercial EFL textbooks, which often present phrasal verbs in grouped sets. Accordingly, the 24 items were divided into four blocks of six verbs, each representing a different particle to avoid overlap that could confound learning (e.g., Block A: bring out, lay down, give in, take up, come off, go over; Block B: make out, sit down, hold back, turn up, cut off, get on).

Each phrasal verb was presented with a learner-friendly definition and example sentence**,** slightly adapted from the PHaVE List in consultation with two Saudi English instructors to ensure cultural and contextual relevance. A lexical frequency check using VocabProfile via Compleat Lexical Tutor (www.lextutor.ca) showed that all component words fell within the first 3000 most frequent words in English: 95.4% at the 1000 level, 99.1% at the 2000 level, and 100% by the 3000 level. Given the learners' proficiency profiles, the individual components of the phrasal verbs were expected to be familiar, allowing focus on the phrasal constructions themselves.

To avoid testing effects or priming from a pretest, a separate norming group was recruited from a more advanced cohort at the same universities. These students, all in their final year

of study, completed the same cued recall test planned for the main study's posttest. They were given immediate feedback after the test to support their learning. The results revealed negligible prior knowledge of the selected phrasal verbs, confirming the suitability of the items for the main sample. This norming procedure aligns with established precedents in vocabulary research (e.g., Boers et al., 2017; Ferguson, Siyanova-Chanturia, & Leeming, 2021; Strong & Boers, 2019a, 2019b), where unfamiliarity with the target items is a prerequisite for assessing intervention effects.

4.3. Research Design

This study adopted a within-subjects experimental design, whereby each participant engaged with two instructional conditions: trial and error and copy-based learning, across a set of 24 phrasal verbs. The phrasal verbs were divided into four balanced blocks, and the sequence in which learners encountered the conditions and verb blocks was counterbalanced to minimize order effects.

The experimental tasks were administered via Qualtrics (Qualtrics, Provo, UT), which was accessed through the students' smartphones. This platform enabled automatic recording of learners' typed responses, while also standardizing time-on-task across conditions. The study was conducted across two classroom sessions during regular English lessons.

- **Session one** involved the two instructional treatments (trial and error, copy), interspersed with distractor tasks designed to clear working memory before testing, followed immediately by the posttests for each condition.
- **Session two**, held one week later, included only the delayed posttests.

Students completed all tasks in their usual classroom environment, and the experimental activities were framed as part of their regular vocabulary learning tasks. While participants were informed about the immediate posttests in the first session, they were not made aware of the delayed posttests in advance. This decision was made to prevent intentional preparation or rehearsal that could influence retention scores, thereby preserving the ecological validity of the results.

Due to ethical and curricular constraints at Saudi public universities, it was not feasible to include a control group who only completed posttests without receiving instruction. Such a condition would have required removing learners from scheduled lessons without offering educational benefit. Moreover, the students were not English majors and reported minimal English exposure outside the classroom, suggesting that incidental learning was unlikely to significantly affect outcomes between testing sessions. Notably, a decline in scores from immediate to delayed posttests suggested that minimal retention occurred beyond the experimental learning activities, helping to mitigate concerns that the first posttest may have inflated performance on the second. These instructional tasks were designed to maintain ecological validity by aligning with real-world classroom activities, thereby supporting task authenticity within the EFL context.

4.4. Procedure

Participants were informed that they would engage in a vocabulary learning activity focused on English phrasal verbs under two different instructional conditions. In the trial-and-error condition, learners initially encountered gap-fill tasks. These included sentences where the particles of the phrasal verbs were omitted and replaced with underlined blanks. The verbs were presented in boldface, and each sentence was accompanied by a paraphrased definition of the full phrasal verb. Learners were asked to supply the missing particle in a textbox beneath the sentence, even if they had to make an educated guess (see Figure1).

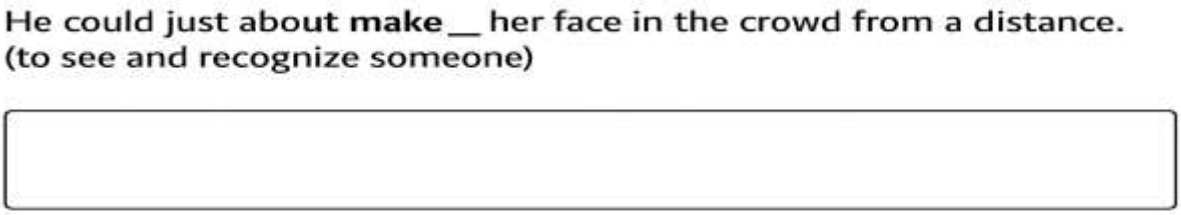

FIGURE 1. Gap-fill exercise in the trial-and-error condition.

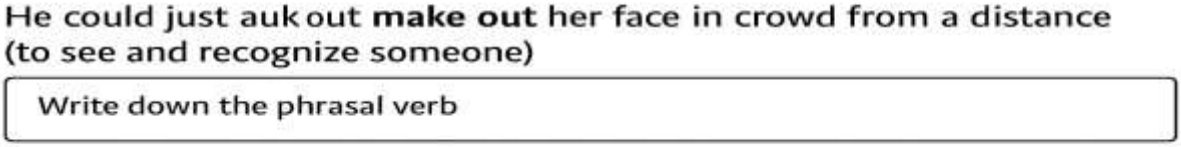

FIGURE 2. Feedback component in the trial-and-error condition.

After four minutes, the screen was updated to reveal the correct answers. The blank spaces were now filled with the correct particles, shown in bold. Students were then instructed to carefully review the complete phrasal verbs and copy them into the provided textboxes as a way of reinforcing their learning (see Figure 2).

To reduce the effects of short-term memory, students were then assigned a distractor task involving the viewing of a 10-minute educational video related to their coursework. Following this, they completed posttests designed to measure retention of that block of phrasal verbs. This process was repeated for the remaining sets of phrasal verbs.

In the copying condition, students were shown sentences containing intact phrasal verbs highlighted in bold. A paraphrased explanation of each verb was displayed alongside the sentence. They were given four minutes to review and memorize the examples (see Figure 3).

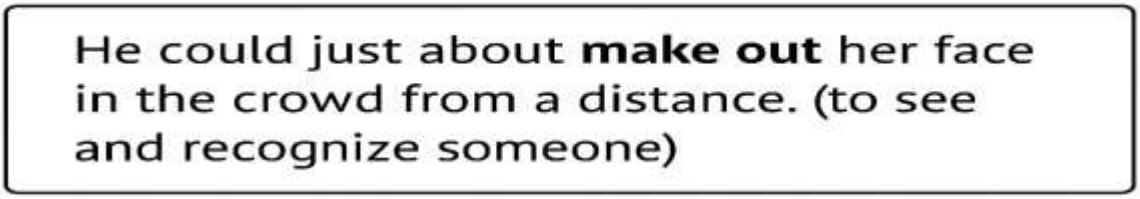

FIGURE 3. The study component in the copy condition.

After the study period, the phrasal verb particles were removed from the sentences and replaced with underlined spaces. The complete phrasal verb remained visible next to the definition. Participants were then instructed to supply the correct particle in a textbox below the sentence (see Figure 4). The same distractor task and posttest procedures followed this condition, and the process was repeated for all sets.

As with the first condition, participants then completed a distractor task followed by posttests. This process was also repeated for the remaining phrasal verb sets.

4.5. Posttests

The effects of the learning conditions were measured using two types of tests: a cued-recall and a multiple-choice test. The cued-recall test was used to measure productive knowledge of the form-meaning connection of phrasal verbs. As shown in Figure 5, participants were presented with the same sentences they encountered in the learning conditions, except the verbs and particles were removed and replaced by two underlined spaces. Participants were asked to recall the entire phrasal verb because the ability to use a phrasal verb entails

knowledge of the whole phrasal verb and not just one individual word (Strong & Boers, 2019b). In addition, for items studied in the guessing condition, participants were asked to provide their original guesses.

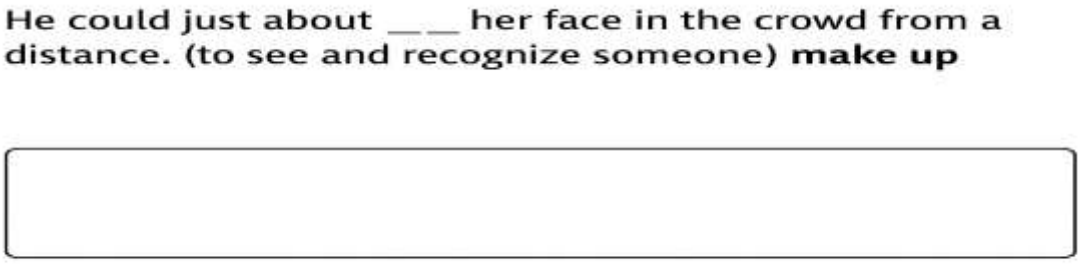

FIGURE 4. Gap-fill exercise in the copy condition.

To evaluate the effects of the two learning conditions, students completed two types of assessments: a cued-recall test and a multiple-choice test.

The cued-recall test assessed learners' productive knowledge of the form-meaning connection of the target phrasal verbs. In this test (see Figure 5), the same sentences used in the learning phase were presented, but both the verb and particle components of the phrasal verb were omitted and replaced with two underlined blanks. Participants were required to reconstruct the full phrasal verb based on their memory, as successful usage demands recall of the entire expression rather than isolated components. For items previously encountered in the guessing condition, learners were also asked to recall and write their original guesses.

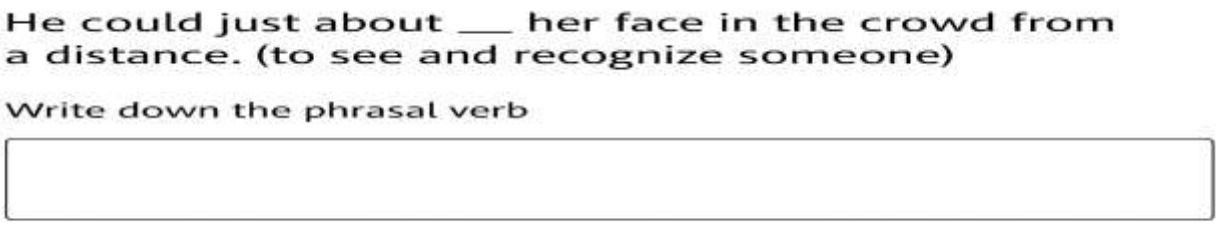

FIGURE 5. Cued-recall test.

In the multiple-choice test (see Figure 6), participants saw the same sentences as in the learning phase, again with the phrasal verbs removed. They were presented with four options: the correct phrasal verb and three distractors. The first distractor shared the verb but used a different particle; the second contained a different verb but retained the correct particle; the third differed in both components. This format tested whether learners had internalized not just form but the precise verb-particle pairing. As with the cued-recall test, students from the guessing condition were asked to recall their initial guesses.

To prevent repetition bias, half of the phrasal verbs were tested in the cued-recall format and the other half in the multiple-choice format. One week later, both tests were administered again to measure retention and vocabulary attrition. Importantly, the test items mirrored the practice sentences, reinforcing the expectation that learners reproduce what they had studied. Students were aware of this objective in advance.

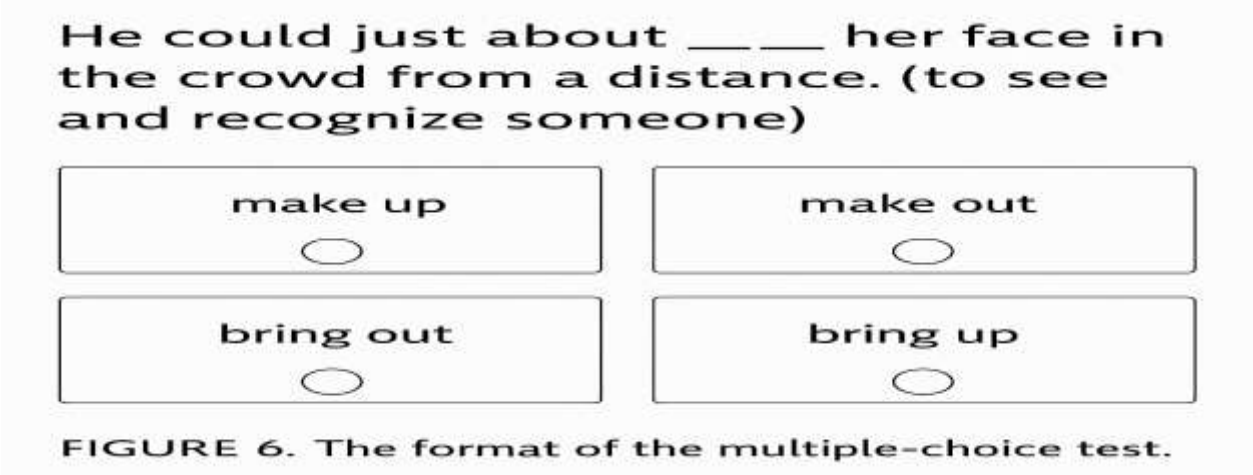

FIGURE 6. The format of the multiple-choice test.

4.6. Data Analysis

Participants' responses to both learning tasks and posttests were evaluated using a binary scoring system (correct/incorrect). Responses that featured an incorrect particle (e.g., using "in" instead of "on") were marked as wrong, as it would have been difficult to distinguish between genuine lexical errors and simple spelling mistakes. However, slight spelling errors in verbs (e.g., "bling" for "bring") were accepted, provided they did not impede identification of the intended target verb. Additionally, responses from the learning activities were analyzed to verify differences in performance between the copying condition and the trial-and-error condition.

Posttest scores were analyzed using mixed-effects logistic regression in jamovi (v2.0, GAMLj module), with condition and retention interval as fixed effects. The analysis included two fixed factors: the type of learning condition (copying vs. trial and error) and the retention interval (immediate vs. delayed). Initially, participants' English proficiency scores and class groupings were entered as fixed predictors, but they were excluded from the final models due to lack of improvement in model fit. Random effects were assigned to both individual participants and test items.

Exercise responses were also analyzed to compare performance across treatment types. The cued-recall and multiple-choice posttest data were analyzed separately using mixed-effects logistic regression models, conducted in jamovi (version 2.0) with the GAMLj module (Gallucci, 2019). The key predictors in the analysis were the type of learning condition (copying vs. trial and error) and the retention interval (immediate vs. delayed). Initially, participants' TOEIC scores and class groups were included as fixed effects in the models, but they were later excluded as they did not enhance model fit. Random effects were accounted for by including both individual participants and test items in the models. In the guessing condition, learners' original responses were captured through open-text entries in Qualtrics. While these entries were stored for verification of recall accuracy, no qualitative analysis (e.g., categorizing error types) was performed, as the study focused on quantitative recall outcomes.

5. Results

5.1. Performance on the Exercise

As anticipated, participants in the copy condition performed considerably better than those in the trial-and-error condition during the gap fill exercise. In the copy group, every response was accurate, resulting in a perfect score of 100 percent. In contrast, only 10.3 percent of responses were correct in the trial-and-error group. This sharp contrast indicates that the way the exercise was carried out had a clear influence on the learners' immediate performance.

It is important to note, however, that the low success rate in the trial-and-error group does not necessarily mean that participants were unfamiliar with the phrasal verbs. The items

used in the exercise contained a limited set of particles, which raises the possibility that some correct answers were the result of chance rather than actual knowledge. Similar findings were reported in previous studies by Strong and Boers in 2019.

Table 1. Mean Percentage of Correct Responses in the Cued-Recall Posttests

	Immediate Posttest		**Delayed Posttest**	
	Mean	SD	Mean	SD
Copy	39%	49%	17%	37%
Trial and Error	31%	45%	16%	37%

5.2. Performance on the Cued-Recall Posttests

Table 1 presents the average accuracy rates from the cued-recall posttests across both instructional conditions. Participants in the copy condition outperformed the trial-and-error group in the immediate posttest (39% vs. 31%). However, this advantage diminished in the delayed posttest, where both groups scored similarly (17% vs. 16%). Mixed-effects logistic regression (Table 2) confirmed significant main effects for learning condition (Cohen's d = 0.34) and retention interval (Cohen's d = 0.72). The copy group significantly outperformed trial-and-error immediately after treatment ($z = 3.245$, $p < .007$), but no significant difference was observed after one week ($z = 0.631$, $p > .05$). Although statistically significant, the effect size (d = 0.27) for the immediate test was modest, suggesting limited educational impact. In contrast, the retention interval effect indicates a notable learning decline, underscoring the importance of reinforcing instruction over time. Knowledge attrition was significant in both groups (copy: d = 0.86; trial and error: d = 0.67), but the rate of forgetting was not statistically different between them. This answers the first research question regarding short- and long-term retention across learning conditions.

Table 2: Fixed Effects Parameter Estimates for the Cued-Recall Test and Retention Interval

	Estimate	SE	Exp (B)	95% Exp (B) CI Lower	95% Exp (B) CI Upper	z	p
Intercept	-1.451	0.140	0.234	0.178	0.308	-10.40	<.001
Learning condition	0.238	0.114	1.328	1.061	1.661	2.48	.013
Retention interval	-1.263	0.118	0.283	0.224	0.356	-10.68	<.001
Learning condition × Retention interval	-0.337	0.228	0.714	0.456	1.117	-1.48	.140

Note. Random effect standard deviations were 1.302 for participants and 0.749 for items. Conditional r^2 value was 0.46.

5.3. Performance on the Multiple-Choice Posttests

Table 3 shows that learners in the copy group slightly outperformed the trial-and-error group in both immediate (84% vs. 81%) and delayed (68% vs. 66%) posttests. These patterns are visually represented in Figure 2, which displays the mean scores alongside their 95% confidence intervals, offering a clearer view of the comparative trends across conditions. As with the cued-recall results, both groups exhibited a decline in performance

over time. However, Table 4 indicates that the learning condition did not significantly predict posttest performance, while the retention interval had a statistically significant effect (Cohen's d = 0.57). Follow-up analyses confirmed this decline in both conditions (copy: d = 0.59; trial-and-error: d = 0.54). These results directly address the second research question, revealing that although both instructional methods were initially effective, neither was sufficient to prevent vocabulary loss over time. While the decay in retention was statistically significant, the differences between instructional methods were educationally minimal.

Figure 2. Mean scores with 95% confidence intervals for the trial-and-error and copying instructional conditions. Bars represent overall average performance across tasks, with error bars denoting 95% confidence intervals.

Table 3: Mean Percentage of Correct Responses in the Multiple-Choice Posttests

	Immediate posttest Mean	Immediate posttest SD	Delayed posttest Mean	Delayed posttest SD
Copy	84%	37%	68%	47%
Trial and error	81%	40%	66%	48%

Table 4: Fixed Effects Parameter Estimates for the Multiple-Choice Test and Retention Interval

	Estimate	SE	Exp(B)	95% Exp(B) Lower	95% Exp(B) Upper	z
Intercept	1.394	0.217	4.029	2.631	6.170	6.408
Learning condition	0.180	0.112	1.197	0.961	1.492	1.608
Retention interval	-1.029	0.114	0.357	0.286	0.447	-9.026
Learning condition x Retention interval	-0.098	0.225	0.906	0.584	1.408	-0.438

.662

Note. Random effect standard deviations were 0.795 for participants and 0.811 for items. Conditional r2 value was 0.32.

5.4. Role of Memory in Recalling Initial Guesses

Table 5 below illustrates how effectively participants remembered their initial guesses during the immediate posttests, and how this memory related to their success in retrieving or recognizing the correct phrasal verbs. The results offer a clear pattern: participants who could recall what they originally guessed were substantially more likely to also retrieve the correct phrasal verb. For instance, in the cued-recall test, 74% of those who remembered their guesses went on to recall the correct target. This trend was even more pronounced in the multiple-choice test, where 98% of participants who recalled their original guesses correctly identified the phrasal verb.

Interestingly, even when learners failed to recall or recognize the correct phrasal verb, many still remembered their initial guesses. As shown in Table 5, in the cued-recall condition, 27% of incorrect responses were accompanied by accurate recall of the earlier guess and in the multiple-choice test, 87% of incorrect answers included successful memory of the original guess.

Table 6 presents the mixed-effects model for the cued-recall task, confirming that memory for the initial guess was a strong and significant predictor of correct responses (Cohen's $d = 1.374$, $z = 6.028$, $p < .001$). Similarly, Table 7 summarizes the regression model for the multiple-choice test, where the predictive effect of remembering a guess also remained statistically robust (Cohen's $d = 1.190$, $z = 2.08$, $p = .038$).

These findings highlight the cognitive impact of making and remembering a guess, even when the guess is incorrect, underscoring the value of active retrieval processes in learning phrasal verbs. Together, these insights address the third research question and deepen our understanding of how initial recall attempts shape subsequent learning outcomes.

Overall, the consistent drop in scores across both cued-recall and multiple-choice tasks underscores the importance of repeated exposure and retrieval to support long-term vocabulary retention.

Table 5: Mean Percentage of Correctly Retrieved Guesses in the Immediate Cued-Recall and Multiple-Choice Posttests

	Correct test answer Mean	Incorrect test answer Mean	Correct test answer Mean	Incorrect test answer Mean
Immediate cued-recall test	74%	27%	98%	87%
Immediate multiple-choice test	42%	44%	13%	34%

Table 6: Fixed Effects Parameter Estimates for Correctly Retrieved Guesses in the Immediate Cued-Recall Test

Parameter	Estimate	SE	Exp(B)	95% Exp(B) Confidence Interval		z	p
				Lower	Upper		
Intercept	–0.0195	0.293	0.981	0.553	1.740	– 0.066	

Parameter	Estimate	SE	Exp(B)	95% Exp(B) Confidence Interval	z	p
Retrieved guess	2.493	0.414	12.094	5.377	27.200	6.028

Note. Random effect standard deviations were 1.024 for participants and 0.706 for items. Conditional R^2 value was 0.42.

Table 7: Fixed Effects Parameter Estimates for Correctly Retrieved Guesses in the Immediate Multiple-Choice Test

Parameter	Estimate	SE	Exp(B)	95% Exp(B) Confidence Interval		z	p
				Lower	Upper		
Intercept	2.950	0.518	19.04	6.90	52.50	5.69	
Retrieved guess	2.160	1.036	8.65	1.14	65.80	2.08	

Note. Random effect standard deviations were 0.942 for participants and 0.812 for items. Conditional R^2 value was 0.22.

6. Discussion

6.1. RQ1: Immediate and Delayed Recall – Productive Knowledge

This study was motivated by previous research that emphasized how the design of gap fill exercises can influence learners' success in acquiring phrasal verbs. The first research question aimed to compare the effects of copying and trial and error gap fill formats on students' ability to produce phrasal verbs. Results from the immediate cued recall test showed that participants in the copying condition performed better than those in the trial-and-error group. A likely explanation is that students in the copy group had more opportunities to interact with the target phrasal verbs. Although the total time on task was balanced across groups, students in the copy group encountered the target phrases twice, once during learning and again during copying. In contrast, students in the trial-and-error group only saw the complete form during the feedback stage. This resulted in an estimated eight minutes of focused exposure in the copy group, compared to approximately four minutes in the trial-and-error group.

Nevertheless, this initial advantage did not continue over time. In the delayed cued recall test, both groups performed similarly. In fact, the copy group showed a slightly higher forgetting rate, although the difference was not statistically significant. These findings suggest that in the long term, both instructional methods were equally effective in supporting productive knowledge. Also, the kind of interference often linked with trial and error in earlier research did not appear to negatively affect learning in this case.

This outcome echoes findings by Strong and Boers (2019b), who reported similar delayed performance between copy based and retrieval-based formats, underscoring how the benefits of passive study tend to erode over time. Likewise, Boers et al. (2017) emphasized that form focused input alone may support immediate gains, but retrieval effort enhances long term retention.

These results suggest that while copying supports surface level encoding that facilitates immediate recall, trial and error may engage more effortful and meaningful processing. This deeper engagement aligns with the levels of processing framework, which proposes that semantically processed information leads to more durable retention than information encoded at a shallow level (Craik and Tulving, 1975). In this context, the act of generating and reflecting on initial guesses, even when incorrect, appears to strengthen memory

consolidation. This aligns with recent evidence in second language acquisition suggesting that retrieval-based learning and error correction can yield lasting lexical gains.

This interpretation aligns with results from Kornell et al. (2009) and Yan et al. (2020), who found that generating incorrect guesses, when followed by feedback, improves memory by promoting deeper encoding and error correction pathways. These studies highlight how productive failure contributes positively to learning outcomes when reflection is involved.

6.2. RQ2: Recognition Accuracy – Receptive Knowledge

The second research question explored how these two instructional approaches influenced learners' receptive knowledge of phrasal verbs. The multiple-choice tests showed no significant differences between the groups, either immediately after the lesson or after a delay. This result implies that the additional exposure in the copy condition did not provide any clear advantage for receptive learning. One possible reason is that recognizing correct answers from a list of similar alternatives may rely more on surface cues than deep processing. For instance, learners might identify the answer based on the initial letters of the verb or particle.

To reduce this risk, the test included three kinds of distractors: one that shared the same verb but had a different particle, another with a different verb but the same particle, and a third with both elements altered. While learners may have quickly eliminated the third and second options, they still had to distinguish between the correct phrasal verb and a distractor with the same verb. This required a detailed mental representation of the full expression. Overall, the findings show that both instructional methods supported the development of this specific lexical knowledge.

6.3. RQ3: Role of Guess Retrieval

The third research question focused on whether learners' ability to recall their initial guesses during practice contributed to improved posttest performance. The results provided compelling evidence that memory for previous guesses played a significant role in both cued recall and multiple-choice outcomes.

As shown in Tables 5 through 7, participants who remembered their earlier guesses were significantly more likely to retrieve or recognize the correct phrasal verbs on the immediate posttests. In the cued recall condition, 74 percent of correct answers were associated with successful guess retrieval. This effect was even more pronounced in the multiple-choice test, where 98 percent of correct responses co-occurred with accurate memory of the original guess. Mixed-effects regression further confirmed that this relationship was statistically robust, with large effect sizes in both testing formats (Cohen's $d = 1.374$ for cued recall; $d = 1.190$ for multiple choice).

These findings suggest that retrieval of prior attempts even if initially incorrect contributes meaningfully to learning outcomes. This supports theoretical models emphasizing the value of retrieval practice and memory reconsolidation. When learners recall their earlier guesses, they likely reactivate semantic and episodic traces, which in turn facilitates deeper encoding and stronger retention of the correct information. This is consistent with the "testing effect" and growing evidence that productive failure can enhance learning when paired with reflection and corrective feedback (Kornell et al., 2009; Yan et al., 2020).

Importantly, this effect was not limited to correct initial responses. Even when guesses were incorrect, remembering them predicted better performance later. This challenges the notion that errors inherently impede learning and instead aligns with perspectives that treat errors as valuable cognitive events—particularly in second language acquisition, where the process of hypothesizing, receiving feedback, and re-encoding may be as critical as correct exposure.

Taken together, the data point to the importance of metacognitive engagement during learning. Encouraging students to recall and reflect on their guesses can transform simple error-making into a productive learning strategy. Future work may investigate how

deliberate strategies such as prompted self-explanation or error journaling might further amplify these benefits.

6.4. Memory of Initial Errors in Trial-and-Error Learning

The third research question examined how phrasal verb learning could be enhanced within a trial-and-error learning environment. Previous research, such as that by Metcalfe and Huelser, emphasized the benefits of recalling one's initial guesses. Building on this idea, the current study asked whether remembering earlier incorrect responses could actually support later learning. Participants were prompted during the immediate posttest to recall the guesses they had made during the gap fill tasks.

The analysis showed that participants were significantly more likely to recall the correct phrasal verb if they could remember their initial guess. This suggests that memory of earlier effort, even if incorrect, may strengthen later retrieval. Interestingly, some students remembered their original wrong guesses but failed to recall the correct answers. This may indicate that errors can sometimes interfere with retrieval, a pattern consistent with findings from Strong and Boers. Still, the overall data from this study suggest that the benefits of reflecting on initial attempts outweighed the potential risks of error recall.

6.5. Pedagogical and Implications

The findings offer clear guidance for instructional design in EFL contexts, particularly when teaching phrasal verbs. First, educators may benefit from incorporating trial-and-error activities followed by immediate feedback and delayed retrieval tasks. This layered approach appears to support long-term retention better than passive copying alone. Second, teachers should encourage students to reflect on their incorrect guesses through metacognitive prompts or classroom discussion, as remembering one's initial errors was strongly associated with improved performance. Finally, digital platforms such as Quizlet or Anki could be used to reinforce phrasal verb learning over time by integrating spaced repetition and cued-recall formats that align with the memory advantages identified in this study.

6.6. Contributions to Theory and Suggestions for Future Research

The study supports the idea that trial and error learning, when combined with reflective recall, can strengthen memory and promote long-term retention. These findings are consistent with cognitive theories such as the Generation Effect and Episodic Encoding, which emphasize the value of self-generated effort and memory retrieval in vocabulary learning.

This study also identifies areas for future research. Although various exercise types appear in EFL and ESL textbooks, not all have been tested for effectiveness. For instance, it remains unclear whether exercises that focus on the verb element differ in impact from those that emphasize the particle. Most existing studies, including those by Strong and Boers, focus on particle-based tasks, leaving room for further exploration. In addition, while this study showed that recalling initial guesses supports later retrieval, the mental processes behind this effect are still not fully understood. Further investigation is needed to explain how memory for guesses contributes to vocabulary development over time.

Finally, the perceptions of learners themselves deserve attention. Students often have strong opinions about which learning methods are most helpful, but these beliefs do not always match actual learning outcomes. Interviews and self-reports can help researchers understand how preferences relate to performance. As phrasal verbs are essential for fluency and comprehension, this area remains a vital topic for continued investigation in both classroom practice and academic research.

7. Conclusion and Limitations

This study aimed to evaluate how two versions of a gap fill exercise; one based on copying and the other on trial and error contributed to the development of both productive and receptive knowledge of phrasal verbs. The findings from the delayed cued recall test

showed that both approaches helped learners improve their ability to recall and use phrasal verbs. Likewise, the delayed multiple-choice test confirmed that each method supported learners' recognition and understanding of these expressions.

An important aspect of this study was the effort to improve learning through trial and error by asking participants to remember their original guesses during the immediate posttests. The results showed that when learners could recall their initial responses, they were more likely to retrieve the correct phrasal verb. This highlights the value of memory for prior attempts in supporting vocabulary learning.

However, more research is needed to better understand how remembering these earlier guesses helps with the learning and recall of correct answers. Future studies should explore the role of episodic memory in guiding successful learning and whether this can be used to design more effective vocabulary learning tasks. Expanding the sample to include learners from different backgrounds would also strengthen the conclusions and improve the generalizability of the findings.

This study is not without its limitations, and several factors should be considered when interpreting the results. First, no pretest was conducted to measure participants' prior knowledge of the selected phrasal verbs. While this might be seen as a shortcoming, the study's focus was not to compare outcomes with a no treatment group, but rather to evaluate the relative effectiveness of two instructional approaches. To mitigate this, a norming procedure with higher proficiency learners from the same educational context served as a benchmark for estimating baseline knowledge. Additionally, since all participants were intermediate English learners already familiar with the lexical components of the target expressions, the task focused on consolidating known items rather than introducing entirely new vocabulary.

Second, the study design involved some susceptibility to guessing effects. In the trial-and-error condition, the low success rate on initial attempts may reflect random guessing rather than knowledge driven retrieval, especially given the limited set of common particles such as "out", "down", "up". Similarly, the use of a multiple-choice format in the receptive test may have inflated scores due to lucky guesses, potentially overestimating true recognition ability. Although such formats are widely used in second language acquisition research, future work may consider complementing them with open ended formats to better capture receptive depth.

A further methodological limitation lies in the within subjects design, which may have introduced carryover effects between instructional conditions despite counterbalancing. Future studies could adopt a between subjects design to better isolate treatment effects and minimize interaction across conditions.

Finally, generalizability is constrained by the relatively homogeneous sample. All participants shared a similar linguistic and educational background. Expanding future research to include learners from diverse first language backgrounds and applying the same methodology to other multiword expressions such as idioms, collocations, or compound verbs would enhance external validity and broaden pedagogical relevance.

References

Al-Otaibi, G. M. (2019). A cognitive approach to the instruction of phrasal verbs: Rudzka-Ostyn's model. Journal of Language and Education, 5, 10–25. https://doi.org/10.17323/jle.2019.8170

Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). Language Teaching Research, 21, 362–380. https://doi.org/10.1177/1362168816651464

Candry, S., Elgort, I., Deconinck, J., & Eyckmans, J. (2017). Word writing vs. meaning inferencing in contextualized L2 vocabulary learning: Assessing the effect of

different vocabulary learning strategies. The Canadian Modern Language Review, 73, 293–318. https://doi.org/10.3138/cmlr.3688
Condon, N. (2008). How cognitive linguistic motivations influence the learning of phrasal verbs. In F. Boers & S. Lindstromberg (Eds.), Cognitive linguistic approaches to teaching vocabulary and phraseology (pp. 133–158). Mouton de Gruyter. https://doi.org/10.1515/9783110199161.2.133
Cornell, A. (1985). Realistic goals in the teaching and learning of phrasal verbs. International Review of Applied Linguistics in Language Teaching, 23, 269–280. https://doi.org/10.1515/iral.1985.23.1-4.269
Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. Studies in Second Language Acquisition, 7, 73–79. https://doi.org/10.1017/S0272263100005167
Darwin, C., & Gray, L. (1999). Going after the phrasal verb: An alternative approach to classification. TESOL Quarterly, 33, 65–83. https://doi.org/10.2307/3588191
Ferguson, P., Siyanova-Chanturia, A., & Leeming, P. (2021). Impact of exercise format and repetition on learning verb–noun collocations. Language Teaching Research. https://doi.org/10.1177/13621688211038091
Gallucci, M. (2019). GAMLj: General analyses for linear models [Jamovi]. https://gamlj.github.io
Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. TESOL Quarterly, 41, 339–359. https://doi.org/10.1002/j.1545-7249.2007.tb00062.x
Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. Language Teaching Research, 19(6), 645–666.
Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? Memory & Cognition, 40, 505–513. https://doi.org/10.3758/s13421-011-0174-0
Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39, 290–296. https://doi.org/10.1037/a0028468
Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? Studies in Second Language Acquisition, 11, 241–255. https://doi.org/10.1017/S0272263100008123
Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 989–998. https://doi.org/10.1037/a0015729
Kovecses, Z., & Szabo, P. (1996). Idioms: A view from cognitive semantics. Applied Linguistics, 17, 326–355. https://doi.org/10.1093/applin/17.3.326
Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1–L2 difference, L1–L2 similarity, or L2 complexity? Studies in Second Language Acquisition, 15, 35–48. https://doi.org/10.1017/S0272263100011657
Liao, Y., & Fukuya, Y. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. Language Learning, 54, 193–226. https://doi.org/10.1111/j.1467-9922.2004.00254.x
Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. TESOL Quarterly, 45, 661–688. https://doi.org/10.5054/tq.2011.247707

Liu, D., & Myers, D. (2020). The most-common phrasal verbs with their key meanings for spoken and academic written English: A corpus analysis. Language Teaching Research, 24, 403–424. https://doi.org/10.1177/1362168818798384

McCarthy, M., & O'Dell, F. (2017). English phrasal verbs in use (2nd ed.). Cambridge University Press.

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia, 138*, 107296. https://doi.org/10.1016/j.neuropsychologia.2019.107296

R. (2014). The benefit of generating errors during learning. Journal of Experimental Psychology: General, 143(2), 644–667.

Potts, R., Davies, G., & Shanks, D. (2019). The benefit of generating errors during learning: What is the locus of the effect? Journal of Experimental Psychology: Learning, Memory, and Cognition, 45, 1023–1041. https://doi.org/10.1037/xlm0000637

Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), A taste for corpora (pp. 173–208). John Benjamins. https://doi.org/10.1075/scl.45.12sch

Seabrooke, T., Hollins, T. J., Kent, C., Wills, A. J., & Mitchell, C. J. (2019). Learning from failure: Errorful generation improves memory for items, not associations. Journal of Memory and Language, 104, 70–82. https://doi.org/10.1016/j.jml.2018.10.001

Side, R. (1990). Phrasal verbs: Sorting them out. ELT Journal, 44, 144–152. https://doi.org/10.1093/elt/44.2.144

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs one-word verbs. International Review of Applied Linguistics in Language Teaching, 45, 109–139. https://doi.org/10.1515/IRAL.2007.005

Skinner, B. F. (1958). Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. Science, 128, 969–977. https://doi.org/10.1126/science.128.3330.969

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 4(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Sonbul, S., El-Dakhs, D. A., & Al-Otaibi, H. (2020). Productive versus receptive L2 knowledge of polysemous phrasal verbs: A comparison of determining factors. System, 95, 102361. https://doi.org/10.1016/j.system.2020.102361

Strong, B. (2013). A cognitive semantic approach to L2 learning of phrasal verbs. The Language Teacher, 37, 28–31. https://doi.org/10.37546/JALTTLT37.5-9

Strong, B., & Boers, F. (2019a). The error in trial and error: Exercises on phrasal verbs. TESOL Quarterly, 53(1), 289–319.

Strong, B., & Boers, F. (2019b). Weighing up exercises on phrasal verbs: Retrieval versus trial-and-error practices. The Modern Language Journal, 103(2), 562–579.

Thomas, M. H., & Dieter, J. N. (1987). The positive effects of writing practice on integration of foreign words in memory. Journal of Educational Psychology, 79, 249–253. https://doi.org/10.1037/0022-0663.79.3.249

Warmington, M., & Hitch, G. J. (2014). Enhancing the learning of new words using an errorless learning procedure: Evidence from typical adults. Memory, 22(5), 582–594.

Webb, S., & Piasecki, A. (2018). Re-examining the effects of word writing on vocabulary learning. International Journal of Applied Linguistics, 169, 72–94. https://doi.org/10.1075/itl.00007.web

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. TESOL Quarterly, 44, 250–273.

Acknowledgements
The author would like to thank the participating students and instructors for their engagement and feedback, as well as the university administration for supporting classroom access and resources. Special thanks to the research assistants who contributed to the data collection and transcription process.

Declaration of Competing Interest
The author declares that there is no conflict of interest.

Ethics Statement
Ethical approval for this study was obtained from the Institutional Review Board of the University of Tabuk. All participants provided informed consent prior to their involvement in the research. Anonymity and confidentiality were ensured throughout the data collection and analysis process, and participants were free to withdraw at any time without consequence.

Appendices: Appendix A: Examples of Trial-and-Error Exercise Items

1. I _____ _____ smoking last year.
2. She _____ _____ her coat and hung it on a hook.
3. They _____ _____ the lights before leaving.
4. He _____ _____ a new way to solve the problem.
5. We _____ _____ our old car for a new one.
6. I _____ _____ all the milk, so we need to buy more.
7. The company _____ _____ the new product last week.
8. She _____ _____ the invitation to the party.
9. He _____ _____ the application form carefully.
10. They _____ _____ the project due to budget issues.

Appendix B: Examples of Copy Exercise Items

1. I gave up smoking last year.
2. She took off her coat and hung it on a hook.
3. They turned off the lights before leaving.
4. He figured out a new way to solve the problem.
5. We traded in our old car for a new one.
6. I used up all the milk, so we need to buy more.
7. The company rolled out the new product last week.
8. She turned down the invitation to the party.
9. He filled out the application form carefully.
10. They called off the project due to budget issues.