Acquiring Collocations Through Multimodal Flashcards

Jessie Takeda
Tohoku University
Emily MacFarlane
Tohoku Gakuin University

Reference Data:

Takeda, J., & MacFarlane, E. (2026). Acquiring collocations through multimodal flashcards. In B. Lacy, M. Swanson, & H. Yoder (Eds.), *LanguageS: Learning, Teaching, Assessing – JALT 50 Years – Challenges and Perspectives*. JALT. https://doi.org/10.37546/JALTPCP2025-22

Collocations are commonly used, non-idiomatic sequences of words that tend to be sequenced together such as “reach a consensus” or “be widely adopted.” Many studies have signified the prevalence of these chunks of language in everyday English (e.g., Goulart, 2019), suggesting their importance in English learning. While incidental learning may be helpful, recent studies have shown that explicit language learning, such as that through multimodal flashcards (Nakata, 2020; Takeda, 2023), can be effective for truly knowing a word (Nation, 2020). Correlational analysis of data from 114 university students revealed interesting patterns in how different practice modes impact short-term versus long-term collocation retention. While writing-based practice showed initial advantages (beta= .365, *p*= .01, relative weight = 27.85%), our statistical analysis uncovered unexpected results about the role of contextualized listening (beta= .235, *p*= .012, relative weight = 30.46%) for sustained vocabulary development. Ultimately, this highlights a significant shift in vocabulary acquisition strategies, demonstrating that receptive auditory input, when paired with rich sentence context, can be more powerful for long-term retention than traditional output-driven tasks.

コロケーションとは、「reach a consensus (合意に達する)」や「be widely adopted (広く採用される)」のように、慣用句ではないが一緒に使われることが多い語の組み合わせである。多くの研究が、これらの語句が日常的な英語で頻出することを示しており (Goulart, 2019)、英語学習における重要性を示唆している。偶発的学習も有益だが、最近の研究では、マルチモーダルなフラッシュカードによる明示的学習の効果が判明している(Nakata, 2020; Takeda, 2023;Nation, 2020) 。114名の大学生のデータの相関分析により、練習方法がコロケーションの短期・長期の定着に与える影響について興味深いパターンが明らかになった。書き取り練習は短期的には優位性を示したが、文脈化されたリスニングが持続的な語彙力向上に果たす役割について予想外の結果が得られた。結論として、豊かな文脈を伴う受容的な聴覚インプットが、従来のアウトプット主導型タスクよりも長期的定着に強力な効果を持ち得ることが示された。

Collocations, at their core, are defined as groups of two or more words that are conventionally and habitually used together by native speakers to create a specific meaning (e.g., Ellis, 1996; Goulart, 2019). The defining characteristic of a collocation is not strict grammatical necessity but rather a sense of naturalness or correctness established through repeated use within a particular context. While alternative word combinations are sometimes grammatically permissible, they often sound awkward, unnatural, or simply “wrong” to the native ear. A classic and illustrative example is the verb-noun pairing *make a mistake*. Although the phrase *do a mistake* may be a common error among non-native speakers, it violates the conventional pairing despite being grammatical. Similarly, while *large* and *big* are synonymous and can generally be substituted for one another, one might speak of a *big decision*, but never *a large decision* (FutureLearn, n.d.). This phenomenon highlights that collocations operate as multiword, unified lexical units, where the choice of one word strongly restricts the choice of the words that surround it. Understanding these conventional pairings is therefore a critical component of achieving fluency and a native-like command of the language.

Quantitative linguistic research has provided compelling evidence of the high frequency of collocations, suggesting that they are fundamental to natural language. A study by Shin and Nation (2007), which used a 10-million-word spoken corpus, revealed that the most common collocations, such as *you know*, *of course*, and *in fact*, occur as frequently as many single words. Further research by Erman and Warren (2009) even suggested that prefabricated language, including collocations, can account for up to half of spoken and written communication. This extensive prevalence indicates that native

speakers do not construct sentences like pearls on a string, and instead, retrieve and utilize larger, prefabricated chunks of language. Learning and mastering these chunks can significantly help second language learners reduce hesitancy and processing time while improving fluency (Du et al., 2022). Therefore, in light of these acquisitional benefits, there has been a growing focus in English Second Language (ESL) and English Foreign Language (EFL) classrooms on the study of multi-word vocabulary units (e.g., Simpson-Vlach & Ellis, 2010; Boers, 2021).

Various methods for learning vocabulary have been practiced over the years. While incidental learning can play a large role (Lee & Pulido, 2017), and undoubtedly occurs due to the frequency of collocations in natural speech, vocabulary words can also be explicitly learned from prepared materials (Kim, 2019; Koh, 2020) such as fill-in-the-blank worksheets (Strong & Boers, 2019; Strong & Leeming, 2024) and flashcards (Nakata, 2020; Spring & Takeda, 2024). One of the benefits of flashcards, particularly online-based flashcards, is that they can be multimodal. Written multiple choice questions (Takeda, 2023), fill-in-the-blank writing exercises (Strong & Leeming, 2024), speaking and listening exercises (Spring & Takeda, 2023), or visual video examples (Nakamura et al., 2024; Kavanagh, 2024) may all be utilized collectively to help language learners to truly know a word (Nation, 2020). This study seeks to determine the effectiveness of studying collocations through multimodal flashcards to see which mode might have the largest effect in the short and long term.

Literature Review

Flashcards are a set of paper or digital cards that are used for studying two pieces of information, or a 'pair', such as a second language (L2) word and its meaning, whether that is the first language (L1) translation, an L2 definition or a synonym (Nakata, 2020). A meta-analysis of 22 studies showed that the use of flashcards leads to moderately large increases in students' proficiency in form-meaning connections compared to other intentional vocabulary-learning activities such as writing or fill-in-the-blanks (Webb et al., 2020). However, traditional flashcards that focus solely on form-meaning connections may be limited when it comes to developing deeper vocabulary knowledge. Recent research on memory indicates that learning is most effective when retrieval practice is paired with various types of contextual cues rather than repeatedly using the same prompts across time (Butowska-Buczyńska et al., 2024). When applying this to flashcard-based vocabulary learning it suggests that studying the target language through multiple modes (e.g. writing, speaking, listening, matching) may be more effective than practicing a single activity type repeatedly. Supporting this approach, Laufer (2016) suggested that learners are more likely to remember a word if they pay attention to multiple details related to the word (pronunciation, meaning, relation to other words, etc.). Furthermore, productive tasks were found to be more effective than receptive ones (Teng & Xu, 2022), highlighting the value of including both receptive and productive modes when studying vocabulary.

Research on flashcard use has explored learning outcomes across different vocabulary types. Xodabande et al. (2022) found that students who used digital flashcards to study discipline specific academic vocabulary performed better than those who used paper flashcards or a standard word list. When using multimodal flashcards to study derivational vocabulary knowledge (word parts), Spring and Takeda (2023) found that using the flashcards generally correlated with higher quiz scores and the writing mode was the most effective in that study. The same researchers looked at using multimodal flashcards for studying phrasal verbs and idiomatic expressions (2024) and the results suggested that while productive modes (speaking and writing) were helpful for short-term memorization, register-specific modes (speaking and listening) were the most significant for long-term retention.

Although flashcard use for collocations remains an understudied area compared to other vocabulary types, broader research on collocation learning has already provided key insights into effective learning conditions. A study by Breslaw and Laufer (2026) looked at the difference between studying collocations grouped thematically versus unrelated and use of L2 only compared to L1/L2 use. They found that both these conditions influenced learning outcomes, with students who studied unrelated collocations using L1/L2 performing the best. When examining what types of tasks were best for learning collocations, Laufer and Breslaw (2025, as cited in Breslaw & Laufer, 2026) suggested that output activities that involved using the target collocation were the most effective. Despite the fact that flashcards have proved effective for other vocabulary types and the recognized importance of learning collocations, no studies have looked at their multimodal use for collocation study. Therefore, this study investigates the effectiveness of different modes for learning collocations within a multimedia online flashcard system, aiming to understand how specific modes influence learning outcomes. Specifically, this study addresses the following research questions:

RQ1. Which flashcard practice modes (reading, writing, speaking, and listening) have the greatest impact on collocation retention in the short-term?

RQ2. Which flashcard practice modes (reading, writing, speaking, and listening) have the greatest impact on collocation retention in the long-term?

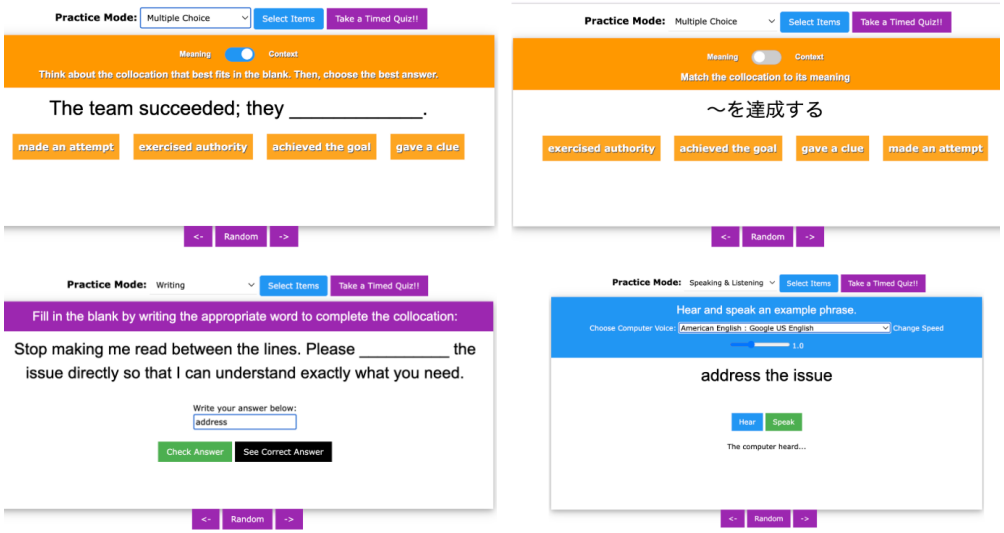
Methodology

A total of 114 first-year EFL students at a national university in Japan participated in this study as part of a university designated general English education class and were taught by one of the two authors. To ensure ethical standards, participation in this study was completely voluntary, informed consent was obtained from all students, and participants were assured that their decision to participate would have no bearing on their academic grades. Scores from a TOEFL ITP® test of general academic English proficiency taken days after this study was completed indicate that most participants fall within the CEFR B1 level (R = 440-650, Mdn = 527, SD = 33.64). Over the course of seven weeks within a reading and writing class, students were tasked with memorizing over 150 academic collocations from the course textbook, *Pathways to Academic English, 4th Edition, v2* (Spring & Scura, 2024), which were originally selected from the *Longman Collocations Dictionary and Thesaurus* (Pearson Education Limited, 2013). The collocations were organized alphabetically by 73 headwords (e.g., *achieve*) with their collocative families (i.e., *achieve a goal*, *achieve an objective*, *achieve an outcome*) and were divided into seven sets to be studied and quizzed on weekly. While incidental experience with the vocabulary items did occur naturally in the classroom and purposefully within worksheets and homework assignments, an in-house-created multimodal flashcard system was developed and utilized by students to explicitly learn the vocabulary. This system is a custom HTML/Javascript tool hosted on a dedicated course website with login access for all enrolled students. The flashcard technology was based on Spring and Takeda (2023) and an example is shown in Figure 1. It offered five study modes for learning collocative vocabulary.

1. **L1/L2 Matching Flashcards:** This mode presented L1 (Japanese language) words and required learners to match them with the correct L2 (English language) collocation from four delayed multiple-choice options. The delay was included because of research by van den Broek et al. (2023) which showed that a short pause between the prompt and potential choices improves language acquisition.
2. **Context Matching Flashcards:** These cards showed a sentence with a blank space. Learners had to select the correct collocation from four delayed multiple-choice options to complete the sentence.
3. **Writing Flashcards:** Like the context matching mode above, this format used a fill-in-the-blank sentence. Following Strong and Boers (2019), learners were required to type the missing collocation.
4. **Listening Flashcards:** This mode displayed the target collocation and allowed students to listen to it in context. Learners clicked a "hear" button to listen to the contextualized sentence, which was read by their device's text-to-speech technology.
5. **Speaking Flashcards:** Students clicked a "speak" button to record themselves saying the phrase. The device's automatic speech recognition (ASR) analyzed the recording to check if the correct collocation was included, and if so, the utterance was marked as "correct." As the ASR was performed by the students' own devices, its accuracy in recognizing variations of English accents were beyond the control of the researchers (see Discussion).

Figure 1

Collocations Flashcard Examples



Students signed into the website with individualized login credentials so that their usage could be tracked by teachers for homework and grading purposes. However, to ensure student privacy, data was later anonymized. Each L1/L2 matching click, context matching click, writing attempt, listening click, and speaking click was recorded weekly. For this study, and as part of seven weekly homework assignments, students were asked to complete at least 10 flashcards of each of the five modes from the weekly set, due at the start of class. Students' data (usage with the website) was downloaded just before corresponding weekly vocabulary quizzes, and homework scores were given based on their completion of the assigned flashcards.

While flashcard usage was assigned as homework, corresponding quizzes were administered through Google Forms under teacher supervision during class time. The seven quizzes were cumulative, meaning that collocations from previous weeks appeared randomly within subsequent quizzes. Additionally, to underscore the importance of studying the vocabulary in various contexts, the weekly quizzes were also multimodal (Uchihara, 2022; Spring & Takeda, 2023). Specifically, each quiz contained three question modes containing between five and ten questions each:

1. **Listening questions:** students listened to five sentences and chose the collocation that they heard from four options (multiple-choice)
2. **Context matching questions:** students read ten sentences containing a blank and chose the collocation that best completed the sentence from five options (multiple-choice)
3. **Writing questions:** students read five sentences containing part of a collocation and filled in the missing piece (written, free answer)

For this study, researchers used three primary metrics to analyze the data. First, the number of clicks on the multimodal flashcards was used to quantify the amount of practice each student undertook in every mode. Second, students' memorization of the collocations was measured by their scores on weekly quizzes. Lastly, a TOEFL ITP® test administered directly after the study was used to balance the quiz data with the students' overall language proficiency. A series of statistical analyses, including correlation analysis calculated in Microsoft Excel, multiple regression analysis using Spring's (2022) online statistics tool, and random forests using Mizumoto's (2023) R-based web application aimed to examine how the use of multimodal flashcards impacted quiz performances over time.

Results

Descriptive statistics for multimodal flashcard usage are shown in Table 1. Though most students did at least some practice with the online flashcards, and despite it being a required homework assignment with student efforts being reflected in their overall grade, some students did not utilize all vocabulary practice modes equally. Specifically, context matching and writing flashcards, both of which provided practice of contextualizing vocabulary in sentences, were the most used practice modes. Listening flashcards, on the other hand, saw half as much overall usage.

Table 1

Descriptive Statistics for Overall Flashcard Clicks by Practice Mode

	Matching L1/L2	Matching Context	Writing	Listening	Speaking
Range	0 - 805	70 - 1078	1 - 1228	0 - 507	0 - 703
Average	171.19	221.16	226.80	150.73	181.61
Median	130	167.50	162	132.50	146.50
S.D.	135.97	154.79	191.07	86.02	118.82

Average scores of quizzes and TOEFL ITP® scores are shown in Table 2. With an average score of 92%, the weekly quizzes were likely considered untroubling to many students, with the listening section often being the easiest section to earn points and the writing section being the most difficult.

Table 2

Descriptive Statistics for Quiz and TOEFL ITP® Scores

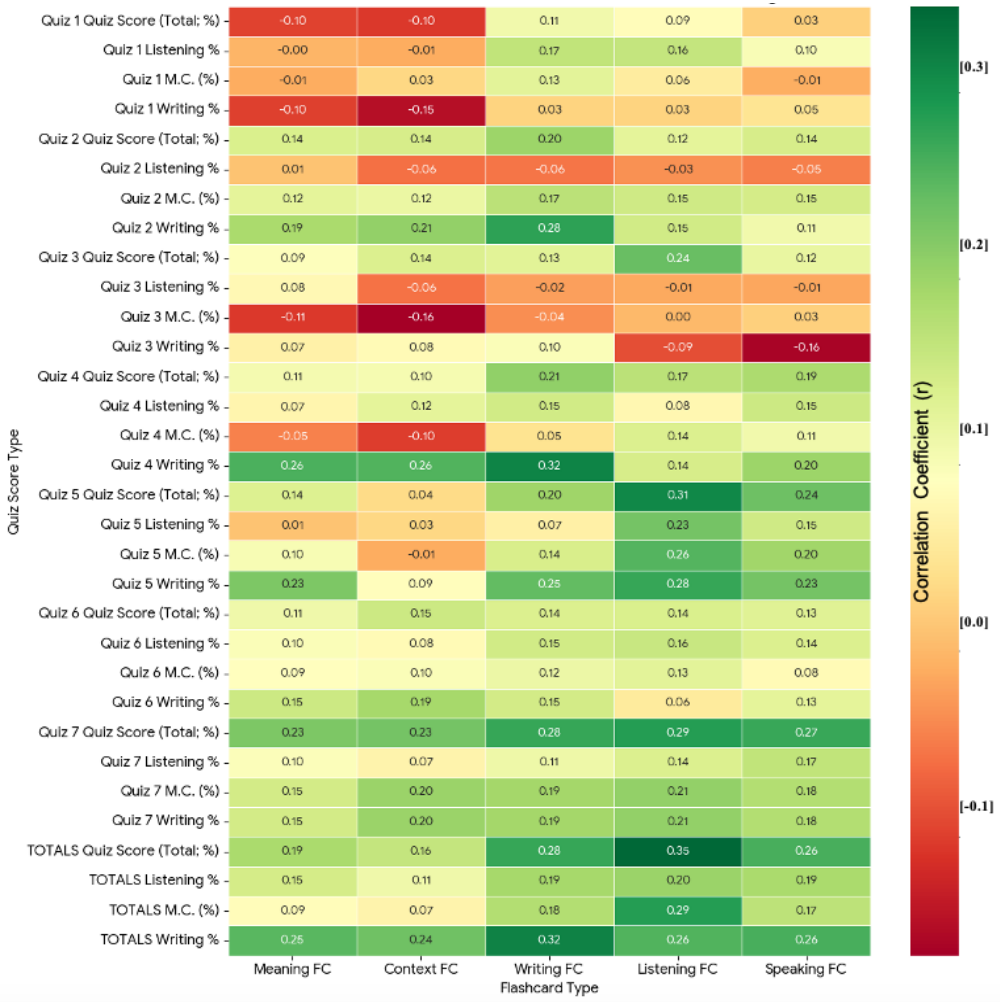
	Overall	Listening	Multiple Choice	Writing	TOEFL ITP®
Range	65 - 100%	80 - 100%	73 - 100%	28 - 100%	440-650
Average	92%	96%	93%	82%	530.83
Median	93%	97%	93%	84%	527
S.D.	6.08	0.04	0.06	0.11	33.64

Correlation matrices were created to determine the relationship between multimodal collocation vocabulary quiz scores and multimodal flashcard usage. By overlaying a heatmap on the correlation matrix, the relationships are seen more clearly. Lighter colors indicate weaker relationships, while red shows negative correlations and green shows positive correlations. As shown in Figure 2, the flashcards had small or even negative effects during the initial weeks of the study. Notably, the usage of writing flashcards began having the greatest positive impacts on quiz scores, particularly, and unsurprisingly on the writing questions. The importance of practicing collocations in the writing mode continued throughout the whole seven weeks, with the correlation coefficients of writing continuing to show some of the highest positive values in the entire matrix. Interestingly, the only flashcard mode with greater correlation than writing was listening. By week 5, usage of listening flashcards seemed to have the greatest overall effect, not only on overall scores, but for all quiz modes. It's worth noting that L1/L2 matching-based flashcards had strong negative correlations with quiz scores, particularly in the first weeks of the study. On the other hand, students who utilized the listening and writing flashcards tended to have higher total quiz scores, listening scores, and writing scores across all quizzes.

Multiple regression analyses are shown descriptively in Table 3 and visually as random forests in Figure 3. TOEFL ITP® scores from an exam taken days after this study concluded serve as a counterweight of overall general English ability. Corroborating the correlative data, multiple regression analysis shows that practicing with writing flashcards was the greatest predictor of vocabulary test scores initially, such as in week 1 (beta= .365, p= .01, relative weight = 27.85%). Interestingly, the relative importance of studying collocations with writing flashcards was similar to that of TOEFL ITP® test scores, indicating that the usage of writing flashcards was as effective as a predictor of performance on vocabulary tests as general English ability. Additionally, in a result unique to week 1 and not reflected in the significance levels of the correlation matrix above, the multiple regression analysis shows that students who studied with both the L1/L2 meaning and context matching-based flashcards tended to score higher on the vocabulary tests. This marked difference between the correlation matrix and multiple regression analysis, along with the negative beta scores, may indicate that there is a learning plateau (i.e., studying with matching flashcards is effective, but beyond a specific point may no longer lead to positive scores). As weeks progressed, however, such as in week 3 (beta= .215, p= .003, relative weight = 44.69%) and week 5 (beta= .235, p= .012, relative weight = 30.46%), students practicing more with listening flashcards tended to score higher that those who practiced less. In week 3, in particular, studying with

Figure 2

Correlation Matrix Heat Map of Multimodal Quiz Scores and Flashcard Usage



listening flashcards was more predictive of the weekly quiz score than TOEFL ITP® test scores, indicating that practicing with listening flashcards alone could make up for a lack of general English ability.

Table 3

Regression Model to Predict Weekly Quiz Scores

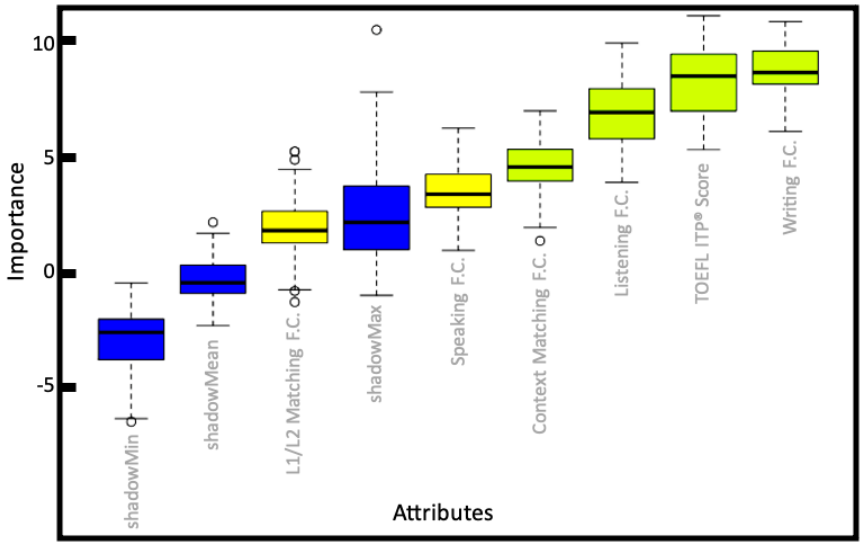
	β Beta	t-value	p-value	Relative Importance	Random Forests
Quiz 1					
TOEFL ITP® Score	0.17	1.821	0.071	29.77%	tentative
L1/L2 Matching Clicks	-0.257	-1.829	0.07	16.30%	confirmed
Context Matching Clicks	-0.243	-1.73	0.086	16.40%	confirmed
Writing Attempts	0.365	2.632	0.01	27.85%	confirmed
Listening Clicks	0.131	1.034	0.304	8.44%	confirmed
Speaking Clicks	0.036	0.29	0.772	1.59%	tentative
Quiz 2					
TOEFL ITP® Score	0.185	1.965	0.052	35.26%	rejected
L1/L2 Matching Clicks	0.044	0.302	0.764	7.40%	tentative
Context Matching Clicks	-0.059	-0.343	0.732	7.95%	rejected
Writing Attempts	0.21	1.576	0.118	30.20%	confirmed
Listening Clicks	0.091	0.817	0.416	10.33%	rejected
Speaking Clicks	0.046	0.404	0.687	9.51%	rejected
Quiz 3					
TOEFL ITP® Score	0.195	2.012	0.047	30.75%	tentative
L1/L2 Matching Clicks	0.018	0.152	0.88	3.21%	tentative
Context Matching Clicks	0.116	0.744	0.458	9.75%	rejected
Writing Attempts	0.029	0.188	0.851	6.35%	rejected
Listening Clicks	0.215	2.193	0.03	44.69%	confirmed
Speaking Clicks	-0.005	-0.05	0.96	4.87%	confirmed
Quiz 4					
TOEFL ITP® Score	0.272	2.912	0.004	52.21%	confirmed
L1/L2 Matching Clicks	0.055	0.432	0.666	3.39%	tentative
Context Matching Clicks	-0.035	-0.263	0.793	2.55%	tentative
Writing Attempts	0.17	1.464	0.146	18.61%	confirmed
Listening Clicks	0.109	1.195	0.235	12.59%	rejected
Speaking Clicks	0.079	0.802	0.425	10.89%	rejected
Quiz 5					
TOEFL ITP® Score	0.277	3.178	0.002	38.41%	confirmed
L1/L2 Matching Clicks	0.115	0.824	0.412	4.29%	tentative
Context Matching Clicks	-0.197	-1.539	0.127	4.62%	rejected
Writing Attempts	0.132	1.124	0.264	9.22%	tentative
Listening Clicks	0.235	2.567	0.012	30.46%	confirmed
Speaking Clicks	0.124	1.247	0.215	13.13%	confirmed
Quiz 6					
TOEFL ITP® Score	0.306	3.33	0.001	66.02%	confirmed
L1/L2 Matching Clicks	-0.009	-0.054	0.957	3.09%	rejected
Context Matching Clicks	0.168	1.138	0.258	10.44%	tentative
Writing Attempts	0.012	0.078	0.938	4.65%	rejected
Listening Clicks	0.119	1.305	0.195	11.65%	rejected
Speaking Clicks	0.02	0.185	0.853	4.14%	rejected
Quiz 7					
TOEFL ITP® Score	0.299	3.45	0.001	41.13%	confirmed
L1/L2 Matching Clicks	0.034	0.213	0.832	6.02%	rejected
Context Matching Clicks	-0.015	-0.094	0.925	6.19%	rejected
Writing Attempts	0.144	1.03	0.306	12.66%	confirmed
Listening Clicks	0.146	1.519	0.132	17.77%	confirmed
Speaking Clicks	0.164	1.634	0.105	16.36%	rejected

	β Beta	t-value	p-value	Relative Importance	Random Forests
Average of All Quizzes					
TOEFL ITP® Score	0.304	3.524	0.001	38.69%	confirmed
L1/L2 Matching Clicks	-0.037	-0.228	0.82	3.88%	tentative
Context Matching Clicks	-0.204	-1.08	0.283	4.83%	confirmed
Writing Attempts	0.379	2.049	0.043	15.71%	confirmed
Listening Clicks	0.258	2.916	0.004	27.54%	confirmed
Speaking Clicks	0.093	0.954	0.342	9.43%	tentative

Random forest analysis confirmed that TOEFL ITP® scores and practicing collocation vocabulary with writing-based flashcards are the most predictive attributes for achieving high scores on weekly vocabulary tests. Surprisingly, the third most predictive attribute is practicing with listening-based flashcards. However, a multimodal strategy for learning collocations remains important.

Figure 3

Random Forests for Average of All Quizzes



Discussion

Our findings reveal a complex relationship between different flashcard modalities and vocabulary acquisition, with some results consistent with established literature while others present a compelling new perspective.

In answer to Research Question 1 regarding which flashcard mode seems to be of greatest impact in the short term, writing-based practice showed the strongest correlation in the short term. This initial advantage was unsurprising, as it aligns with ample previous research which highlights the effectiveness of learning through active recall (Karpicke & Roediger, 2008) and productive tasks (Teng & Xu, 2022). The act of producing the target collocation through writing forces a deeper level of cognitive processing than simple recognition-based tasks, which explains its immediate benefit in predicting short-term quiz performance.

However, and in answer to Research Question 2 regarding which flashcard mode had the greatest impact in the long-term, our statistical analysis uncovered an unexpected and significant role for contextualized listening in sustained vocabulary development over time. This finding presents an interesting paradox to most previous research, as listening is traditionally considered an "input" modality, which is often viewed as less potent for direct vocabulary learning than active production tasks. A potential explanation for this lies in the nature of the listening tasks themselves. The flashcards did not isolate the collocation but presented it within a naturalistic sentence context. This rich contextual information, even if passively received, may have provided a more robust and multifaceted memory anchor for the collocation's meaning and usage. This interpretation aligns closely with digital instructional design research by Li et al. (2022), who demonstrated that the deliberate inclusion of sound within software environments functions as a distinct memory anchor during recall, helping users retrieve targeted content and associated linguistic structures. Furthermore, this long-term auditory advantage highlights a critical divergence from the frameworks of Karpicke and Roediger (2008) and Teng and Xu (2022). While these researchers established that active production and testing are paramount for long-term retention, their studies primarily utilized single-word vocabulary units. Because the target vocabulary in this study were multi-word collocations embedded in naturalistic sentences, the auditory element within and the meaningful context may have provided a more durable long-term cognitive anchor. This finding follows other research by Spring & Takeda (2023), underlining the importance of listening in acquiring multi-word vocabulary units. Finally, since the listening mode was the least utilized, it could be that the students who did engage with it were those who were truly committed to the multimodal learning strategy. For these

learners, the auditory input served as a crucial supplement, reinforcing the form and pronunciation of the collocation in a way that the other, more "output"-based modes could not. Auditory input may have mimicked a more authentic language experience, thereby enhancing the long-term memorability of the collocation beyond the scope of a typical flashcard task.

Despite these promising findings, our study has several limitations that warrant consideration. First, the study's duration of seven weeks may not be sufficient to fully capture the nuances of long-term vocabulary retention. Future research could extend the study period to better assess the sustained impact of multimodal flashcards. Second, the participant pool was limited to first-year EFL students at a single university in Japan, which may affect the generalizability of the findings to different learner populations, proficiency levels, and cultural contexts. Additionally, while we tracked the utilization rates of each study mode, our analysis did not consider the qualitative aspects of student engagement with the flashcards. Future studies should consider gathering qualitative data on how students perceive and interact with each mode to gain a deeper understanding of the cognitive processes at play. Moreover, as the ASR component of the Speaking Flashcards was performed on the students' own devices, the researchers had no control over the ASR engine's recognition of Japanese-accented English, which means some students may have experienced more "incorrect" markings, affecting the number of times they practiced that flashcard type. However, as the weekly quizzes did not include a speaking component, students' test performance was not affected and so this would not compromise the validity of the study results. Another limitation of the tool's design is that it did not measure the time taken for students to click. This could be explored in the future as a possible measure of engagement and learning.

The results of our study have important implications for both language teachers and curriculum developers. While traditional production-based activities like writing are valuable for initial learning, our findings suggest that a multimodal approach that purposefully incorporates contextualized listening is crucial for fostering long-term vocabulary retention. Teachers can integrate this by encouraging the use of audio-based features in their learning materials, employing dictation exercises with authentic language, or using online flashcard systems that provide real-world examples in an auditory format. The surprising effectiveness of the listening mode underscores the need to move beyond a singular focus on output and to embrace diverse input modalities as essential components of a robust collocation vocabulary acquisition strategy.

Conclusion

This study investigated the effectiveness of a multimodal digital flashcard system for learning academic collocations among university EFL students over a seven-week period. The empirical findings reveal a complex, dynamic relationship between the usage of multimodal flashcards and collocative vocabulary retention over time. While it is unsurprising that writing-based flashcards provided an immediate short-term advantage through active recall and deeper cognitive processing, contextualized listening emerged as an unexpected driver of sustained, long-term vocabulary development for these multi-word vocabulary units. Even though the listening mode was the least used by students, suggesting that they found little immediate acquisitional benefits, by listening to embedded collocations within naturalistic sentences, the contextualized auditory input seems to have significantly enhanced long-term memorability. Conversely, traditional L1/L2 translation-based matching tasks yielded initial negative correlations, signaling a potential learning plateau inherent to simple recognition-based learning.

These insights carry vital implications for contemporary language pedagogy. To move beyond a singular focus on vocabulary output or rote translation, educators and curriculum developers should embrace diverse, multimodal input strategies. Intentionally incorporating contextualized auditory elements alongside production tasks within digital learning frameworks is crucial for fostering genuine, long-term vocabulary retention. Ultimately, while this study provides a foundational understanding of multimodal collocation acquisition, future research should expand beyond a seven-week timeline, investigate highly diverse learner populations, and incorporate qualitative measures to fully capture the nuances of student engagement and cognitive processing in digital environments.

Bio Data

Jessie Takeda is a teacher of English for academic purposes at Tohoku University in Sendai, Japan, where she researches about and practices multimodal learning. She has served on the handbook team for the JALT National and PanSIG conferences for three years. takeda.jessica.leigh.giddens.e1@tohoku.ac.jp

Emily MacFarlane is an experienced English language educator based at Tohoku Gakuin University, where she teaches general education English. She is also an adjunct lecturer in Academic Reading and Writing at Tohoku University. Her work focuses on improving general education English courses for mixed-major university students, with a particular interest in speaking fluency development. She serves as Program Chair for JALT Sendai. emily.macfarlane@mail.tohoku-gakuin.ac.jp

References

Boers, F. (2021). *Evaluating second language vocabulary and grammar instruction: A synthesis of the research on teaching words, phrases, and patterns.* Routledge.

Breslaw, R., & Laufer, B. (2026). Learning new collocations: The effects of grouping and language of instruction, *System, 136*, 103873, https://doi.org/10.1016/j.system.2025.103873

Butowska-Buczyńska, E., Kliś, P., Zawadzka, K., & Hanczakowski, M. (2024). The role of variable retrieval in effective learning. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(44), e2413511121. https://doi.org/10.1073/pnas.2413511121

Du, X., Afzaal, M., & Al Fadda, H. (2022). Collocation use in EFL learners' writing across multiple language proficiencies: A corpus-driven study. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.752134

Ellis, N.C. (1996) Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*,18(1), 91–126. http://doi.org/10.1017/S0272263100014698

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk, 20*, 29–62. https://www.semanticscholar.org/paper/The-idiom-principle-and-the-open-choice-principle-Erman-Warren/bb9b41f92fb34faa965ad126a10340629276de04

FutureLearn. (n.d.). *What are collocations?* FutureLearn. https://www.futurelearn.com/info/courses/improve-ielts-speaking/0/steps/98854

Goulart, L. (2019). The use of collocations across proficiency levels: A literature review. *BELT (Brazilian English Language Teaching Journal)*, *10*(2), e34129. https://doi.org/10.15448/2178-3640.2019.2.34129

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. http://doi.org/10.1126/science.1152408

Kavanagh, B. (2024). The animated short film as a tool for teaching English for general academic purposes: Integrating inferences and connotations. *ATEM Journal: Teaching English through Multimedia, 29*, 85–97.

Kim, S. (2019). The effects of using images and video clips for vocabulary learning: Single words vs phrasal verbs. *STEM Journal, 20*(2), 19–42. https://doi.org/10.16875/stem.2019.20.2.19

Koh, S. (2020). Initial formulaic sequence approach for EFL learners: Based on film corpus-based patterns. *STEM Journal, 21*(4), 1–17. https://doi.org/10.16875/stem.2020.21.4.1

Laufer, B. (2016). From word parts to full texts: Searching for effective methods of vocabulary learning. *Language Teaching Research*, *21*(1), 5–11. https://doi.org/10.1177/1362168816683118

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research, 21*(1), 118–135. https://doi.org/10.1177/1362168816637381

Li, Y., Finch, S., Sweany, N. W., & Kwok, O. M. (2022). Exploring sound use in embodied interaction to facilitate learning in a digital environment. *The Journal of Applied Instructional Design, 11*(4), 155–169.

Mizumoto, A. (2023). Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests. *Language Learning, 73*(1), 161–196. https://doi.org/10.1111/lang.12518

Nakamura, S., Spring, R., & Sakurai, S. (2024). The impact of ASR-based interactive video activities on speaking skills: Japanese EFL learners' perceptions. *TESL-EJ, 27*(4), 1–18. https://doi.org/10.55593/ej.27108a5

Nakata, T. (2020). Learning words with flash cards and word cards. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 304–319). Routledge. https://doi.org/10.4324/9780429291586

Nation, I. S. P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15–29). Routledge. https://doi.org/10.4324/9780429291586

Pearson Education Limited. (2013). *Longman collocations dictionary and thesaurus*.

Shin, D., & Nation, P. (2007). Beyond single words: The most frequent collocations in spoken English. *ELT Journal, 62*(4), 339–348. https://doi.org/10.1093/elt/ccm091

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Spring, R. (2022) Free, online, multilingual statistics for linguistics and language education researchers. *Center for Culture and Language Education, Tohoku University 2021 Nenpo, 8*, 32–38. https://doi.org/10.13140/RG.2.2.12037.63202

Spring, R., & Scura, V. (Eds.). Institute for Excellence in Higher Education, Tohoku University. (2024). *Pathways to academic English, 4th edition, version 2*. Tohoku University Press.

Spring, R., & Takeda, J. (2023). The effect of multimodal flashcards on L2 derivational vocabulary knowledge: A preliminary analysis of attempts and quiz question modes. *Japan Association of Language Education and Technology Kanto Chapter, 8*, 1–24. https://doi.org/10.24781/letkj.8.0_1

Spring, R., & Takeda, J. (2024). Teaching Phrasal Verbs and Idiomatic Expressions Through Multimodal Flashcards. *Journal of English Teaching through Movies and Media*, *25*(2), 40–53. https://doi.org/10.16875/stem.2024.25.2.40

Strong, B., & Boers, F. (2019). The error in trial and error: Exercises on phrasal verbs. *TESOL Quarterly, 53*(2), 289–319. https://doi.org/10.1002/tesq.478

Strong, B., & Leeming, P. (2024). Evaluating the application of a gap-fill exercise on the learning of phrasal verbs: Do errors help or hinder learning? *TESOL Quarterly, 58*(2), 726–750. https://doi.org/10.1002/tesq.3248

Takeda, J. (2023). Pathways to Academic English: Multi-modal online tools: Initial analysis of advantages and student feedback. *Tohoku University, Institute for Excellence in Higher Education Annual Bulletin, 9*, 43–48. https://tohoku.repo.nii.ac.jp/record/139698/files/2434-8317-2023-9-43.pdf

Teng, M., & Xu, J. (2022). Pushing vocabulary knowledge from receptive to productive mastery: Effects of task type and repetition frequency. *Language Teaching Research*, *29*(2), 588–606. https://doi.org/10.1177/13621688221077028 (Original work published 2025)

Uchihara, T. (2022). How does the test modality of weekly quizzes influence learning the spoken forms of second language vocabulary? *TESOL Quarterly*, *57*(2), 595–617. https://doi.org/10.1002/tesq.3176

van den Broek, G. S. E., Gerritsen, S. L., Oomen, I. T. J., Velthoven, E., van Boxtel, F. H. J., Keter, L., & van Gog, T. (2023). Optimizing multiple-choice questions for retrieval practice: Delayed display of answer alternatives enhances vocabulary learning. *Journal of Educational Psychology*, *115*(8), 1087–1109. https://doi.org/10.1037/edu0000810

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

Xodabande, I., Iravi, Y., Mansouri, B., & Matinparsa, H. (2022). Teaching academic words with digital flashcards: Investigating the effectiveness of mobile-assisted vocabulary learning for university students. *Frontiers in Psychology, 13,* 893821. https://doi.org/10.3389/fpsyg.2022.893821