

What Types of Illustrations are Effective in Intentional Vocabulary Learning? Simple or Complex? Clear or Unclear?

May 2026 – Volume 30, Number 1
<https://doi.org/10.55593/ej.30117a5>

Satoshi Ide

University of Tsukuba, Japan
<s2320051@u.tsukuba.ac.jp>

Akifumi Yanagisawa

University of Tsukuba, Japan
<yanagisawa.akifum.ft@u.tsukuba.ac.jp>

Abstract

Visual aids are commonly used in language learning; however, empirical evidence on their benefits for vocabulary learning is mixed. As the effectiveness of illustrations could depend on their characteristics, the impact of different illustration features (e.g., clarity and complexity—simple vs. detailed images) as well as other potential moderating factors on intentional vocabulary learning was assessed in this study. Two experiments were conducted for this purpose, respectively adopting vocabulary lists and flashcards. In both experiments, Japanese EFL university students learned unfamiliar words in one of three conditions—no-illustration, simple illustrations, or complex illustrations—and their learning was assessed via pretests, immediate posttests, and delayed posttests. The results showed that while no significant effects of illustrations were found for list learning, with flashcards, the effects of illustrations were moderated by word concreteness. Specifically, both simple and complex illustrations facilitated the learning of concrete words, whereas simple illustrations hindered the learning of abstract words.

Keywords: deliberate vocabulary learning, illustration, levels of processing, concreteness, vocabulary size

As vocabulary acquisition is the foundation of language learning (e.g., Nation, 2022; Schmitt, 2010), visual aids have been widely used to facilitate this process (e.g., Tomlinson, 2008; Wright, 1989). The benefits of illustrations as learning aids are supported by a considerable body of psychological and educational research. Referred to as the *picture superiority effect*, it has been established that, as images are more memorable than language (e.g., Paivio et al., 1968; Shepard, 1967), illustrations can enhance memory retention in paired-associate learning (Paivio & Yarmey, 1966). They have been shown to be beneficial for incidental vocabulary learning (e.g., Ramezani et al., 2021, but, see also Boers et al., 2017) and for comprehension and vocabulary learning through online dictionaries (Dziemianko, 2022). Although these

findings suggest that illustrations positively impact both learning outcomes and motivation, studies focusing on deliberate vocabulary learning offer mixed findings.

While some authors find the use of illustrations beneficial for vocabulary learning (e.g., Yowaboot & Sukying, 2022), others report null effects (e.g., Boers et al., 2009). These inconsistencies could potentially arise due to the differences in the illustration features. For example, according to the levels of processing theory (Craik & Lockhart, 1972), the complexity of illustrations (e.g., pictogram vs. realistic image) affects their effectiveness because the complexity changes the amount of information a learner processes when viewing these visual aids. In addition, other illustration characteristics, such as clarity (i.e., how clearly an illustration expresses the meaning of the corresponding target word) and word concreteness (because concrete words may be easily depicted by an illustration), may moderate the effect of illustrations on learning. Thus, by clarifying the effectiveness of various illustration types in intentional vocabulary learning, this study aims to elucidate the role of illustrations as learning aids in vocabulary learning.

Background

Empirical Research on the Use of Illustrations for Vocabulary Learning

The benefits of illustrations for L2 vocabulary learning are supported by a significant body of research (Alsalihi, 2020; Bates & Son, 2020; Yowaboot & Sukying, 2022). For example, Bates & Son (2020) investigated the impact of simple illustrations on vocabulary learning by comparing the performance of learners provided with simple illustrations with that of their counterparts who received no illustrations. Their results indicated that the group aided by simple illustrations scored significantly higher on fill-in-the-blank questions on both immediate posttest and delayed posttest than the group without illustrations. More recently, Yowaboot and Sukying (2022) examined the effectiveness of digital flashcards in vocabulary learning relative to L1 translation only and found them superior in terms of both receptive and productive vocabulary knowledge gains. Similar conclusions were reached by Alsalihi (2020), who investigated the impact of posters including visual images (vs. no posters) on vocabulary learning, and noted a positive effect on meaning recall.

In contrast, using a within-participants design, Boers et al. (2009) found no benefits of illustrations for idiom learning. They compared two conditions, whereby participants received verbal explanations only or explanations supplemented with illustrations. Both conditions completed multiple-choice questions related to the origins of idioms and exhibited comparable performance on the immediate posttest comprising fill-in-the-blank questions. The authors attributed these findings to the possibility that, when both verbal and visual inputs are provided, they competed for the limited attention capacity, which may have led to better retention of the visual input at the cost of verbal input.

Despite not directly examining the effect of adding illustrations to deliberate vocabulary learning, Carpenter and Olson (2012) compared L2 vocabulary learning in the cases of (a) L1 translation words and (b) pictures. In contrast to what would be predicted by the picture superiority effect, their results showed no superiority of pictures over translations in learning and revealed that participants were much more confident in recalling the target words from pictures than from translations. When their confidence bias was removed by warning and by retrieval practice, learning with pictures led to greater gains than learning with translations. These findings suggest that learners' perceptions of effectiveness in learning might influence the actual effectiveness of learning with illustrations.

While illustrations are generally found effective for vocabulary learning, some prior studies suggest otherwise. However, with the exception of Boers et al. (2009), whose participants

engaged in exercises individually in a computer room, extant research largely involves classroom-based approaches. Thus, to rigorously test the effects of illustrations and provide additional evidence on their role in vocabulary acquisition, investigations based on controlled experiments are required.

Theoretical Foundations for Using Illustrations to Enhance Vocabulary Acquisition

We hypothesized that the provision of illustrations would enhance vocabulary learning due to increased retrieval routes (Zheng et al., 2016). Zheng et al. (2016) investigated whether the effect of L2 word learning on long-term memory retention could be increased when studying with increased retrieval routes. In Experiment 1, in the two initial learning sessions, native Chinese speakers studied a list of unknown Japanese words with the corresponding Chinese translations, with each session including a mediator (hint) for each word pair (i.e., a set of a word and a phrase that assists in associating the L1–L2 words; e.g., for 泥棒 [“thief” in Japanese]–小偷 [Chinese] pair, 泥巴 [mud] with 小偷把泥巴糊脸上以防被认出 [the thief plastered mud on the face] was used as a mediator). After a thirty-minute break, participants engaged in receptive retrieval practice, in which they were presented with Japanese words and asked to type in the corresponding Chinese words and the main idea of the mediator sentences. During this retrieval phase, half of the participants practiced with one mediator (using the same mediator), and the rest with two mediators (using two different mediators). In a posttest administered one week later, participants were presented with the Japanese target words and asked to provide the corresponding mediators and Chinese words. The results revealed that the participants who studied in the two-mediators condition achieved greater retention compared to those who practiced with the same mediator twice. In the current study, similarly, when students learn new vocabulary with illustrations, these illustrations would serve as additional mediators to retrieve the meanings of the corresponding target words. In such a case, even when the direct form–meaning mapping is attenuated, the memory of illustrations could assist learners in recalling the meanings of the words.

Empirical Research on Types of Illustrations and Vocabulary Learning

Earlier studies have compared the effectiveness of different types of illustrations. For instance, Mahdi and Gubeily (2018) examined the effect of unusual (or strange, such as a cat barbecuing with a grill) illustrations on vocabulary learning by comparing the posttest performance of a target-words-only, a normal illustration, and a strange illustration group. Their results showed that the group exposed to strange illustrations scored significantly higher on meaning recall in both immediate and delayed posttests compared to the other two groups. Kaplan-Rakowski (2019) also investigated the effect of illustrations on vocabulary learning by comparing a group provided with 2D illustrations and a group provided with 3D illustrations. The fill-in-the-blank test results showed no significant differences between the groups.

Illustration Complexity

While there is a substantial body of research examining the relationship between illustrations and vocabulary learning, illustration complexity (e.g., simple vs. complex) was never directly investigated as a potential moderating factor, to our knowledge. The levels of processing theory (Craik & Lockhart, 1972) posits that information encoded at a deeper semantic level will be retained better in memory. When this framework is applied to the effect of illustrations on memory, it suggests that image complexity affects the amount of semantic information an illustration conveys. For instance, the illustration on the right side of Figure 1 provides more semantic details than the simpler image on the left. Such differences might significantly influence the learning process, especially how effectively an unfamiliar word is processed and retained in memory.

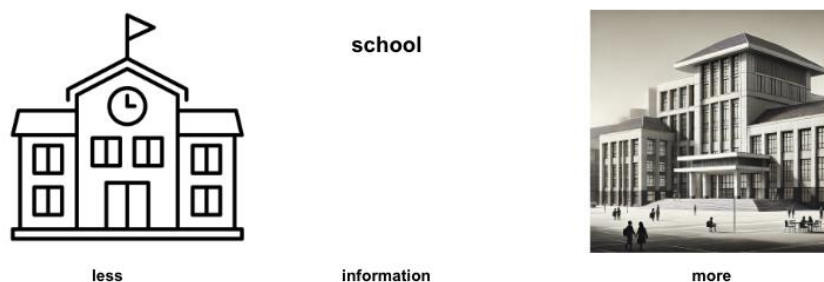


Figure 1. Examples of Simple and Complex Illustrations

In this study, the amount of information an illustration entails was operationalized as complexity—a dichotomous variable denoting an illustration as either simple or complex (see, e.g., Skulmowski and Rey, 2018 for a similar discussion). When assessing relative complexity (simple vs. complex), three factors were considered: the degree of realism, the presence of contextual background information, and the density of lines used in the illustrations. These aspects were deemed relevant, as complex illustrations tend to be more realistic compared to simple illustrations. Simple illustrations often omit details and are more symbolic than realistic, resembling pictograms (Skulmowski & Rey, 2018). Moreover, simple illustrations tend to depict only the target words, while complex illustrations often include additional background information related to the target words (e.g., the depiction of people for the picture of school as in Figure 1). Finally, complex illustrations tend to feature more lines in their compositions to convey greater detail (Snodgrass & Vanderwart, 1980). While we could have operationalized complexity by only manipulating one factor (e.g., different degrees of contextual background information), we decided to focus on relative amounts of information because this is the first study directly investigating this aspect.

Other Factors Potentially Influencing Learning with Illustrations

While numerous factors may potentially influence vocabulary learning with illustrations, we focused on three variables: illustration clarity as an illustration-related variable, word concreteness as a word-related variable, and vocabulary size as a learner-related variable. These variables might facilitate or hinder the effect of illustration provision on deliberate learning and potentially moderate the effects of illustration complexity.

Illustration Clarity. The term “illustration clarity” reflects how clearly the illustration represents the meaning of the word it depicts. While there is little direct empirical research on this topic, illustrations with high levels of clarity presumably facilitate semantic understanding better than those characterized by low clarity. When an illustration communicates the meaning of a word clearly, it may serve as an effective cue to retrieve its meaning. Consequently, when the word is encountered in a posttest, illustrations presented during the learning phase would facilitate the retrieval of its meaning from one’s memory. If the illustration clearly expresses the meaning of the corresponding word, learners may be more likely to recall its meaning. Conversely, if the meaning of an illustration is unclear, the appropriate meaning of the corresponding word would be difficult to be recalled even when the learner remembers the illustration. Furthermore, illustration complexity might also moderate the effect of illustration clarity. For example, complex illustrations might only be effective if they express meaning clearly, whereas the clarity of meaning is less important for simple illustrations.

Word Concreteness. Word concreteness refers to the degree of ease with which a particular word can be visualized. For example, “elk” and “clergyman” are highly concrete because they describe concrete objects, while “affection” and “accusation” are less concrete because they are intangible concepts. When the word is easy to imagine, it is more likely to be retained

(Paivio, 1969; Paivio et al., 1968; see also Boers et al., 2017 for review in the context of L2 learning). The more concrete the word, the more likely it may evoke a sensory image that mediates associative learning (Ellis & Beaton, 1993). As concrete words are easier to convey via illustrations than abstract words, word concreteness is hypothesized to be positively correlated with vocabulary learning. Furthermore, complexity might also play a role in this relationship. Considering simple illustrations might be the most effective for concrete words, whereas complex illustrations could be more suitable for abstract words, which often require more detailed illustrations to convey their meanings.

Another outcome that can be hypothesized is that abstract words benefit more from the provision of illustrations than concrete words. This is based on the idea that concrete words, which can be pictured easily in learners' mind, may evoke a greater degree of visual processing from the word themselves, so adding an illustration may have little value. In contrast, abstract words may benefit more from illustrations because their less imagery provoking nature can be complemented by the provision of illustrations (see Boers et al., 2017 for a similar discussion in the context of multimodal glossing).

Vocabulary Size. For the purposes of the current investigation, vocabulary size is defined as the number of words the learner knows. Empirical evidence indicates that learners with a larger vocabulary (or greater prior vocabulary knowledge) are more likely to learn unknown words than those who have a smaller vocabulary (e.g., Horst et al., 1998; Kasahara & Yanagisawa, 2024; Webb & Chang, 2015). As vocabulary size is associated with learners' proficiency (Jeon & In'nami, 2022), it might influence the effect of illustration complexity, whereby students with smaller vocabulary (or lower proficiency) benefit more from illustrations, as visual aids are primarily used for beginners (Maley, 2011).

The Present Study

As was discussed in the previous sections, most of the evidence in favor of illustrations as effective vocabulary learning aids was not derived from controlled experiments. Furthermore, the roles that characteristics of illustrations and words play in vocabulary learning remain unclear. To address these gaps in pertinent literature, as a part of this study, we investigated (a) the effectiveness of illustrations as intentional vocabulary learning aids and (b) the effects of potentially influential factors (i.e., illustration complexity and clarity, word concreteness, and learners' vocabulary size). We also surveyed learners' perception of effective types of illustrations on vocabulary learning. Accordingly, the following three research questions guided our study:

RQ1: Do illustrations facilitate deliberate vocabulary learning? If so, which type (complex vs. simple illustrations) is most effective?

RQ2: To what extent are illustration clarity, word concreteness, and vocabulary size related to the illustration effect on vocabulary learning?

RQ3: What illustrations do learners consider effective in vocabulary learning?

To answer these questions, two experiments featuring different learning materials were conducted. Vocabulary lists were used in Experiment 1, while flashcards were adopted for Experiment 2 after considering the results obtained from Experiment 1. Vocabulary list learning was chosen as the initial method of deliberate vocabulary learning because of its frequent use alongside word card learning (Nakata, 2008; Webb et al., 2020).

Experiment 1: Vocabulary List

Participants

The initial cohort for Experiment 1 comprised 97 EFL learners at a university in Japan ranging in age from 18 to 20 ($M = 19$). All participants were L1 Japanese speakers and majored in several fields such as humanities, information science, and biology. The data pertaining to 12 students who did not participate in the learning sessions and 8 students for whom the assigned learning condition was unidentifiable were excluded from the analysis of vocabulary learning. As we also excluded data related to one participant who fell asleep during the experiment, a total of 76 participants' data remained. However, we included the responses from the post-experiment questionnaire (e.g., their rating of illustration clarity and their perception on vocabulary learning with illustrations). Only one participant had lived in an English-speaking region for longer than one year. In accordance with the between-subjects design adopted for this study, the participants were randomly assigned to one of the three conditions: no-illustration (target words + L1 translations), simple illustration (target words + L1 translations + simple illustrations), and complex illustration (target words + L1 translations + complex illustrations).

Learning Material: Word List

The word list used in Experiment 1 featured 20 words selected from A New Corpus-based English for General Academic Purposes Wordlist for Advanced Learners at Japanese universities, named BABILON2000 (Ishikawa, 2018). Low-frequency words from the highest level (level 9) of BABILON2000 were selected as presumably unfamiliar vocabulary for most of the participants. Fifteen out of twenty Japanese translations of the target words were also adopted from BABILON2000, and the translations of the remaining five words—as their translations in BABILON2000 were not suitable for the current experiment—were taken from a Japanese–English dictionary (Inoue & Akano, 2003).

Both simple and complex illustrations expressing the meanings of target words were created using ChatGPT-4 (OpenAI, 2024), resulting in 40 items in total. For example, to create a simple illustration for *elk*, ChatGPT-4 was provided with the following prompt: “Create a simple illustration that can be drawn in 10 seconds to represent ‘elk’.” Because ChatGPT-4 tends to create complex illustrations by default, “Create a complex illustration to represent ‘elk’” was used for the complex alternative. Both authors evaluated several generated illustrations and chose those that expressed the meaning clearly and were suitable for each complexity condition.

As participants were assigned to three learning conditions, separate word lists were created for each group. For the no-illustration condition, target words alone were presented next to their corresponding L1 translations. For the simple-illustration and complex-illustration conditions, the target words were presented alongside their corresponding illustrations (either simple or complex) and L1 translations. The lists were printed on A4 paper (see the Supplementary Materials Appendix A for the word lists used in the experiment).

Questionnaire

To obtain participants' background information and data on their vocabulary learning experience and perceptions, a questionnaire was administered via Google Forms. In addition to providing basic information, the participants were required to select the types of illustrations they perceived as effective for vocabulary learning from the following options: simple illustrations, complex illustrations, self-drawn pictures, strange illustrations, photographs, none of the above, and others (multiple selections were allowed).

Assessment Measures

Vocabulary Knowledge of Target Words. To account for participants' prior knowledge of the target words, a meaning recall test was conducted as a pretest before the learning session. To measure vocabulary knowledge development, the same meaning recall test as well as a meaning recognition test were conducted after the learning session in this order. The meaning recall test involved translation of each target word into Japanese, whereas the meaning recognition test required that participants select one corresponding Japanese translation from four choices. As in Waring and Takaki (2003), for distractors of the meaning recognition tests, we used L1 words that were semantically related and of the same part of speech as the correct answer (see the Supplementary Materials Appendix A for the entire tests used in the experiment). We did not administer the meaning-recognition test as a pretest in order to minimize the influence on the ultimate results of learning from the pretest (e.g., by guessing the right answer through eliminating unlikely options on multiple-choice questions) (see also Webb, 2024, for the discussion of testing effects). The order of words was randomized for each test. All tests were administered in paper and pencil format. The participants were explicitly told to study the target words but not informed about the immediate or delayed posttests.

Other Variables that Potentially Influence Vocabulary Learning with Illustrations: Illustration Clarity, Word Concreteness, and Vocabulary Size. Vocabulary Size Test for Japanese EFL Learners Using the New JACET List of 8,000 (Hamada et al., 2021) was used to measure participants' vocabulary size. Considering the proficiency level of the university students that took part in this experiment, we used the last six sections (levels 3–8) from eight levels. This test includes 120 multiple-choice questions regarding the meaning of English words and participants are required to select one English word from four options that in their view matches the Japanese word presented in each question. Participants' response was converted into a proportion and used as a proxy of their breadth of vocabulary knowledge (Anderson & Freebody, 1981).

The word concreteness ratings were adopted from Brysbaert et al. (2014), who reported concreteness ratings for English words based on responses from more than 4,000 individuals using a 5-point Likert scale.

Illustration clarity—how clearly the meaning of the word was conveyed through the illustration—was evaluated by the participants on a 5-point Likert scale, ranging from 1 = “not clear at all” to 5 = “very clear.”

Procedure

Experiment 1 was conducted as a part of academic English presentation skills classes, where students learn how to read academic journal articles and make presentations. Participation in the language learning experiment was a part of the coursework and was designed to help students experience and understand how quantitative experimental research can be conducted. After the experiment, the students were informed of the tentative results and discussed the potential interpretations and caveats regarding the findings. Prior to commencing the experiment, its purpose was explained, along with the methods used for data collection and analysis, and the students' role as participants. Once they had the opportunity to ask any questions, students were provided with the consent form. They were informed that they could decide whether or not their data would be included in the analysis and were assured that their decision would not influence their class grade or result in any negative consequences. Once the participants signed the consent forms, they completed the demographic questionnaire, followed by a meaning recall test as a pretest. Next, participants were given five minutes to learn the 20 target words under one of the three randomly assigned conditions [The learning time was

determined on the basis of a small pilot study with three Japanese graduate students in English language education, in which they were given 7 or 10 minutes to study the target words. To minimize ceiling and floor effects in at least one of the two test formats (meaning-recall and -recognition), we ultimately decided on a study time of 5 minutes]. During the learning session, participants could only use the distributed material and were banned from using other equipment. Once the five minutes had lapsed, participants were presented with an irrelevant 15-minute-long TED talk video as a distractor aimed at alleviating the recency effect. The experiment concluded with the meaning recall and meaning recognition tests. This order of tests was chosen to avoid the meaning recognition test—where the L1 translations for each target words were presented—influencing the meaning recall test performance. Two weeks later, the same tests were repeated, and a questionnaire on the participants' perception of vocabulary learning with illustrations and their evaluation of clarity of illustrations used in this experiment was administered.

Scoring and Analysis

For both meaning recall and meaning recognition tests, one point was awarded for each correct response. To answer RQs 1 and 2, we employed generalized linear mixed-effects models (GLMMs) using the *lme4* (Bates et al., 2015) and *lmerTest* packages (Kuznetsova et al., 2017) in R Studio (version 2024.09.0+375). Specifically, the accuracy on the immediate posttest and delayed posttest was used as the dependent variable, while learning condition (i.e., no-illustration, simple illustration, vs. complex illustration), vocabulary size, word concreteness, and illustration clarity were treated as independent variables, and their main effects and relevant interactions were included in the model. Pretest response accuracy for meaning recall was included as a covariate in models predicting the accuracy on the meaning recall tests, but not in the models predicting the accuracy on the meaning recognition tests. This was because the influence of pretest responses was not adequately estimated on meaning recognition tests, likely due to the ceiling effects. In all cases, participants and words were included as random intercepts. The two test formats (recall and recognition) and test timings (immediate and delayed) were analyzed separately. The contrast for the learning condition was specified using simple effect coding (Goodman et al., 2025), with the no-illustration condition being the reference. The numerical predictor variables were z-score standardized in case of convergence issue (i.e., with illustration clarity and vocabulary size for Experiment 1). All materials, datasets, analytic code, and supplementary materials are publicly available at OSF (<https://osf.io/p63qt/>).

Results of Experiment 1

RQ1: Do Illustrations Facilitate Vocabulary Learning? If So, Which Type (Complex vs. Simple Illustrations) is Most Effective?

Table 1 shows the average meaning recall test scores for each learning condition, indicating that the differences across conditions were rather small. On the immediate posttest, participants assigned to the no-illustration condition achieved the highest mean scores (no-illustration: 17.12 [85.6%], simple illustration: 16.25 [81.3%], complex illustration: 16.81 [84.1%]). On the delayed posttest, those in the complex-illustration condition attained the highest mean scores (no-illustration: 10.77 [53.9%], simple illustration: 11.11 [55.6%], complex illustration: 11.96 [59.8%]). The omnibus tests (Type II Wald chi-square tests) revealed no significant differences in the meaning recall test performance across conditions for either the immediate posttest or the delayed posttest (see Table 2).

Table 1. Descriptive Statistics for the Meaning Recall Tests

Condition:	No Illustration				Simple Illustration				Complex-illustration			
	<i>n</i>	<i>M</i>	(%)	<i>SD</i>	<i>n</i>	<i>M</i>	(%)	<i>SD</i>	<i>n</i>	<i>M</i>	(%)	<i>SD</i>
Pretest	25	1.88	(9.4)	1.51	24	1.12	(5.6)	1.03	27	1.78	(8.9)	1.28
Immediate Posttest	25	17.12	(85.6)	3.38	24	16.25	(81.3)	3.45	27	16.81	(84.1)	3.26
Delayed Posttest	22	10.77	(53.9)	3.78	18	11.11	(55.6)	2.81	23	11.96	(59.8)	3.84

Note. *n* = Sample size; *M* = Mean; *SD* = Standard deviation. Possible maximum score was 20.

Table 2. Omnibus Test Results for Meaning Recall Tests

Immediate Posttest	Delayed Posttest
$\chi^2(2) = 0.61, p = .739$	$\chi^2(2) = 1.54, p = .463$

The results of the meaning recognition tests are not reported in the text because their scores indicated the ceiling effect (i.e., mean scores exceeded 19 out of 20). For readers interested in the details, see Appendix B in the supplementary materials. Similarly, in the analyses pertaining to RQ2, we focused exclusively on the meaning recall tests because of the ceiling effects observed with data on the meaning recognition tests.

RQ2: To What Extent are Illustration Clarity, Word Concreteness, and Vocabulary Size Related to the Illustration Effect on Vocabulary Learning?

Illustration Clarity. Table 3 presents the detailed results of statistical analyses. GLMM results indicated that illustration clarity had no significant main effect on either immediate posttest scores or delayed posttest scores. Additionally, no significant interaction effect was observed between condition and illustration clarity on either the immediate posttest or the delayed posttest. These findings suggest that illustration clarity was unrelated to the learners' performance on the delayed meaning recall posttest under the vocabulary list condition.

Word Concreteness. Word concreteness did not exhibit a significant main effect on the meaning recall posttest scores irrespective of the test timing. Similarly, no significant interaction effect was observed between condition and word concreteness in either case, immediate posttest. These results suggest no clear association between word concreteness and learning outcomes.

Vocabulary Size. Vocabulary size had a significant main effect on the immediate and the delayed posttest. The interaction effect between condition and vocabulary size approached but did not reach statistical significance on the immediate posttest and was not significant on the delayed posttest. Therefore, larger vocabulary size seems to result in greater vocabulary retention.

Table 3. Summary of Main Effects and Interactions of Potentially Influential Variables on Meaning Recall Posttest Scores

Factor	Test Timing	Main Effect		Interaction Effect	
		Odds Ratio	<i>p</i>	χ^2 (df)	<i>p</i>
Illustration Clarity	Immediate	1.28	.092	$\chi^2(1) = 0.03$.860
	Delayed	1.18	.222	$\chi^2(1) = 1.58$.209
Word Concreteness	Immediate	1.43	.202	$\chi^2(2) = 3.24$.198
	Delayed	0.93	.827	$\chi^2(2) = 1.65$.438
Vocabulary Size	Immediate	1.68	.001	$\chi^2(2) = 5.13$.077
	Delayed	1.31	.037	$\chi^2(2) = 1.08$.583

Note. Illustration Clarity and Vocabulary Size were z-score standardized.

RQ3: What Illustrations do Learners Consider Effective in Vocabulary Learning?

Figure 2 presents the types of illustrations learners considered effective for vocabulary learning. Simple illustrations were rated as the most effective by 58 participants (74.4% of the sample), while only seven (9.0%) opted for strange illustrations, and four (5.13%) chose complex illustrations.

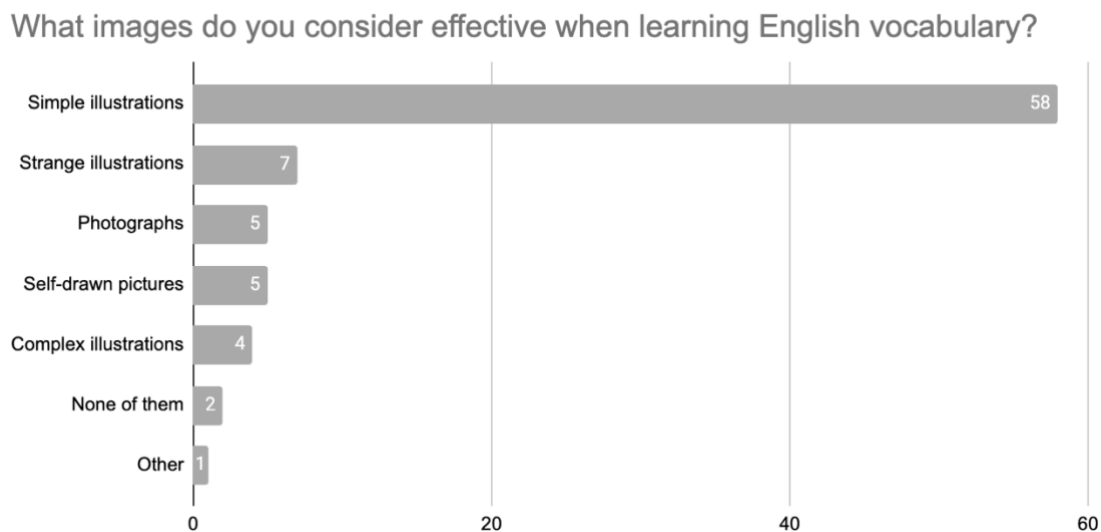


Figure 2. Learners' Perception of Illustration Type Effectiveness for Vocabulary Learning

Note. The results are based on 78 participants' responses. The number presented in each horizontal bar indicates the number of participants that selected the category. Multiple responses were accepted.

Discussion for Experiment 1

RQ1: Do Illustrations Facilitate Vocabulary Learning? If So, Which Type (Complex vs. Simple Illustrations) is Most Effective?

Three learning conditions (no-illustration, simple illustration, and complex illustration) were compared to establish which condition is most suitable for intentional vocabulary learning with a vocabulary list. Given the lack of statistical significance and small mean differences, the

results suggest that presenting illustrations in a vocabulary list may not effectively enhance vocabulary learning regardless of the illustration type (simple vs. complex).

Similar conclusions were reached by Boers et al. (2009), but contrasting findings were reported by Alsalihi (2020) and Yowaboot and Sukying (2022). While Alsalihi (2020) as well as Yowaboot and Sukying (2022) found positive illustration effects, their studies were conducted in classroom settings, and the learning sessions were spread over a few days. Therefore, students that took part in these investigations were exposed to teacher's instructions over a longer period and potentially benefitted from homework-based learning, which might have contributed to the observed illustration effects. The discrepancy between these findings and those obtained in the present study could have also potentially arisen due to the learners' deliberate focus on the target words and L1 translations, while ignoring the corresponding illustrations in this study.

RQ2: To What Extent are Illustration Clarity, Word Concreteness, and Vocabulary Size Related to the Illustration Effect on Vocabulary Learning?

The analysis pertaining to this research question revealed that vocabulary size was the only factor showing a clear relationship with initial learning and retention. Learners with larger vocabulary sizes outperformed those with smaller vocabulary sizes, aligning with the findings obtained in previous studies on incidental and deliberate vocabulary learning (e.g., Horst et al., 1998; Kasahara & Yanagisawa, 2024). However, as no clear relationships were found for learning conditions, word concreteness, or illustration clarity, illustrations may be of limited benefit for vocabulary retention by the current study's participants at any proficiency level, which was relatively high for the university EFL learners.

RQ3: What Illustrations do Learners Consider Effective in Vocabulary Learning?

According to the survey results, learners generally think simple illustrations are more effective than complex illustrations and expressed their perception that illustrations facilitate vocabulary learning. Interestingly, these perceptions contradict the obtained results suggesting a lack of clear advantage of illustrations irrespective of their type. On the other hand, this finding is in line with the observations made in extant L2 research indicating that learners' perception of efficacy in a specific learning strategy may not corroborate the actual learning outcome (e.g., Nakata & Suzuki, 2019).

Experiment 2: Flashcards

Purpose of Experiment 2

Experiment 1 was conducted to investigate the effects of illustrations on intentional vocabulary learning in a controlled setting; however, no clear illustration effects were found regardless of the illustration features. One possible reason for this outcome is that participants might have tended to ignore the illustrations during the learning phase. As word lists were used in this experiment, the question remained whether a different visual aid would help learners process illustrations and potentially cause illustration effects. For example, learning with flashcards, where learners attempt to recall illustrations printed on the other side of the card by being cued with the L2 target words, may decrease the probability of illustrations being ignored while learning. Additionally, flashcards provide opportunities for retrieval practice, which refers to the mental process of accessing stored information and is known to enhance learning outcomes (e.g., Barcroft, 2007). When viewing the target word on one side of the flashcard, learners try to retrieve both the illustration and the L1 translation on the opposite side. As this act of retrieval likely enhances the memory retention of the illustrations, learning with flashcards

would lead to a more pronounced influence of illustrations. Enhanced memory of illustrations may serve as additional cues for learners to retrieve the meaning of corresponding target words.

Another aspect of Experiment 1 that would benefit from revision concerns the difficulty of the target words and the meaning-recognition tests. The learning gains observed in Experiment 1 were relatively high, possibly because of the highly proficient participants, and some of the target words and meaning-recognition tests could have been easier for them. Therefore, incorporating target words that are unfamiliar to a larger proportion of participants, and increasing the difficulty of the meaning recognition test, could enhance the ability to compare learning gains across conditions.

Participants

The sample for Experiment 2 comprised 24 EFL learners from a university in Japan ranging in age from 18 to 28 ($M = 21.5$) [Our sample size was determined by practical constraints related to the data collection period and a common rule of thumb that a minimum sample size of 20–30 participants is generally required for each cell (e.g., see Simmons et al., 2011). Prior to data collection, we decided to recruit at least 20, targeting 30 participants. By the end of the data collection period, data had been obtained from 24 participants, and we concluded the data collection. This decision was based on the idea that for repeated measures ANOVA regarding our main focus, i.e., the comparison between three illustration conditions (with $\alpha = .05$, correlation between measures = .5), median and large effect sizes (Cohen's $f = .25$ and $.40$) would provide 74% and 99% of estimated statistical power, respectively, despite not being directly suitable for the current study, which adopted generalized linear mixed-effects models]. All participants were L1 Japanese speakers and majored in several fields such as humanities, information science, and biology. None of them had experience of living in English-speaking countries.

Learning Material: Flashcards

For this experiment, 21 target words were selected, 17 of which were the same as those used in Experiment 1, while the remaining 4 words—for which the accuracy rate exceeded 20% in the pretest of Experiment 1—were replaced with new words adopted from BABILON2000 (Ishikawa, 2018). This replacement was intended to lower the mean scores on posttests, and thus to facilitate a more meaningful comparison across conditions.

As both complex and simple illustrations were created for each of the 21 words, 42 illustrations were prepared for this experiment. As in Experiment 1, all illustrations were created using ChatGPT-4 (OpenAI, 2024). As shown in Figure 3, in every condition, only the target word was shown on the front. The back contained only the L1 translation for the no-illustration condition, the simple illustration plus the L1 translation for the simple-illustration condition, and the complex illustration plus the L1 translation for the complex-illustration condition.

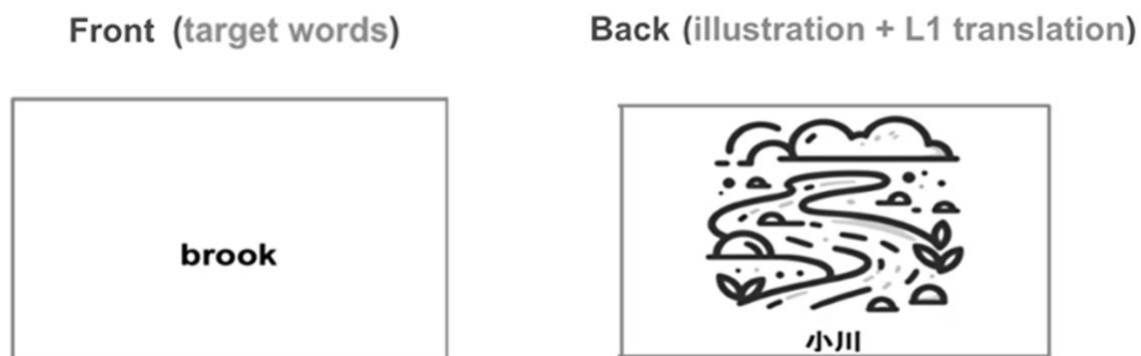


Figure 3. Flashcards as the Learning Material: Simple Illustration Condition

Questionnaires

The same questionnaire of participants' background information and vocabulary learning perception as in Experiment 1 was used. Additionally, the degree to which participants looked at the illustrations while learning was assessed using a 5-point Likert scale question, with one indicating "did not look at the illustrations at all", and five indicating "looked at the illustrations very much".

Assessment Measures

Vocabulary Knowledge of Target Words. While the same method for vocabulary knowledge assessment as in Experiment 1 was utilized, different distractors were adopted in the meaning recognition tests to mitigate the impact of potential ceiling effects by increasing the test difficulty [The meaning recognition tests in Experiment 1 included distractors that were prepared for each test item and were semantically related, while in Experiment 2, correct answers for other words served as distractors. Distractors for each question were randomly selected while ensuring that each distractor was used with equal frequency].

Other Variables that Potentially Influence Vocabulary Learning with Illustrations: Illustration Clarity, Word Concreteness, and Vocabulary Size. The same tests and ratings as in Experiment 1 were adopted for potentially influential variables.

Study Design

While a between-participants design was adopted for Experiment 1, Experiment 2 was based on a within-participants design. This change was made to increase the statistical power across conditions and to accommodate a relatively small sample size. In line with Experiment 1, three learning conditions were employed (i.e., no-illustration, simple illustration, and complex illustration), each condition containing seven words, for a total of 21 words studied. Participants were recruited on a voluntary basis. The order of the conditions and target words assigned for each condition were counterbalanced, and participants were randomly assigned to one of these three counterbalancing conditions.

Procedure

Experiment 2 was conducted outside the normal class schedule and took approximately 90 minutes to complete. In line with Experiment 1, participants were provided with the consent form and were informed that they could decide whether their data would be included in the analysis without any adverse consequences. They were also assured that they could withdraw from the research at any point in the experiment. At the start of the session, all participants completed the questionnaire, followed by the meaning recall test as a pretest and the vocabulary size test, after which they were randomly assigned to one of three counterbalancing conditions in order to learn the target words with flashcards. The learning phase involved three conditions, each lasting 100 seconds, during which participants were told to study a set of seven words using flashcards. In total, participants spent five minutes learning the 21 words. The participants could only rely on the distributed material and were prohibited from using any other equipment. After the learning session, the meaning recall test was conducted followed by the meaning recognition test. The same tests and the questionnaire were repeated two weeks later. Participants were compensated with 2,000 yen for their participation.

Scoring and Analysis

To account for the model identifiability issues—which likely arose as very few target words were known to the participants—we excluded the pretest variable from our models pertaining to all immediate posttest analyses [To evaluate the effect of not including pretest responses as a covariate, a sensitivity analysis was conducted by removing the data points from the posttest

when they were answered accurately on the pretest. The results corroborated the original results with the entire dataset (see RMarkdown files available at the OSF page for the details of the sensitivity analysis)]. Otherwise, the same scoring and analysis methods as in Experiment 1 were adopted. While the inclusion of random slopes is frequently suggested in within-participants research designs (Barr et al., 2013; see also Matuschek et al., 2017), our models with *condition* as the random slope for *participant* indicated a singular fit issue. Moreover, their AIC values favored the random-intercepts models, suggesting that the inclusion of the random slopes did not enhance the model fit (Matuschek et al., 2017). Consequently, we opted for the models including random intercepts only.

Results of Experiment 2

Attention to Illustrations While Learning

The results of the 5-point Likert scale on the degree of attention to illustrations during the learning phase revealed that 12 participants (50.0%) reported that they looked at the illustrations very much, 11 (45.8%) reported that they looked at the illustrations, and one (4.2%) reported "neutral". This indicates that we successfully ensured that the vast majority of the participants directed their attention to the illustrations when learning with the flashcards rather than ignoring them.

RQ1: Do Illustrations Facilitate Vocabulary Learning? If So, Which Type (Complex vs. Simple Illustrations) is Most Effective?

Meaning Recall Test. Table 4 shows the average meaning recall test scores under each learning condition. The complex-illustration group achieved the highest performance on both immediate posttest (no-illustration: 5.12 [73.1%], simple illustration: 5.25 [75.0%], complex illustration: 5.62 [80.3%]) and delayed posttest (no-illustration: 3.08 [44.0%], simple illustration: 3.12 [44.6%], complex illustration: 3.79 [54.1%]). However, no significant differences in the meaning recall scores on the immediate posttest were noted, while those on the delayed posttest approached but did not reach statistical significance (see Table 5). Thus, while the delayed posttest results suggest a potential advantage of complex illustrations for long-term vocabulary retention, no clear difference across learning conditions was observed.

Table 4. Descriptive Statistics for the Meaning Recall Tests

Condition:	No-illustration			Simple-illustration			Complex-illustration		
	<i>M</i>	(%)	<i>SD</i>	<i>M</i>	(%)	<i>SD</i>	<i>M</i>	(%)	<i>SD</i>
Pretest	0.17	(2.4)	0.38	0.08	(1.1)	0.28	0.17	(2.4)	0.38
Immediate Post	5.12	(73.1)	1.42	5.25	(75.0)	1.62	5.62	(80.3)	1.53
Delayed post	3.08	(44.0)	1.25	3.12	(44.6)	1.36	3.79	(54.1)	1.59

Note. Total sample size (*N*) = 24; *M* = Mean; *SD* = Standard deviation. Possible maximum score was 7.

Table 5. Omnibus Test Results for Meaning Recall Tests

Immediate Posttest	Delayed Posttest
$\chi^2(2) = 3.14, p = .208$	$\chi^2(2) = 5.07, p = .079$

As in Experiment 1, the results of meaning recognition tests indicated the ceiling effect (i.e., mean scores exceeded 6 out of 7). Detailed results are reported in the supplementary materials, Appendix B. Accordingly, in the analyses of RQ2, we focus exclusively on the meaning recall tests to provide a more targeted explanation for the obtained findings.

RQ2: To What Extent are Illustration Clarity, Word Concreteness, and Vocabulary Size Related to the Illustration Effect on Vocabulary Learning?

Illustration Clarity. Table 6 presents the details of the statistical analyses. As in Experiment 1, only data from the simple-illustration and complex-illustration conditions were analyzed for illustration clarity. For the meaning recall immediate posttest performance, the effect of illustration clarity did not reach significance on either immediate posttest or delayed posttest. No clear interaction effect between condition and illustration clarity was found on either posttest.

Table 6. Summary of Main Effects and Interactions of potentially influential variables on Meaning Recall Posttest Scores

Factor	Test Timing	Main Effect		Interaction Effect	
		Odds Ratio	<i>p</i>	χ^2 (df)	<i>p</i>
Illustration Clarity	Immediate	1.51	.113	$\chi^2(1) = 1.57$.210
	Delayed	1.21	.419	$\chi^2(1) = 0.02$.901
Word Concreteness	Immediate	1.85	.007	$\chi^2(2) = 1.78$.410
	Delayed	1.59	.060	$\chi^2(2) = 7.59$.023
Vocabulary Size	Immediate	1.29	.278	$\chi^2(2) = 2.00$.367
	Delayed	1.08	.594	$\chi^2(2) = 3.77$.152

Note. Vocabulary Size was z-score standardized.

Word Concreteness. The meaning recall immediate posttest results showed that word concreteness had a significant effect on the participants' scores, suggesting that the more concrete the words, the greater the learning gains. A similar trend—albeit not statistically significant—was observed for the delayed posttest.

As for the interaction effect between condition and word concreteness, a significant interaction was observed for the delayed posttest (see Table 7 presenting the detailed results of simple main effect analysis), while no interaction was observed for the immediate posttest. These results suggest that the illustration effect differed based on the word concreteness (see Figure 4 illustrating the relationship between delayed posttest scores and word concreteness across the three learning conditions). While the estimated retention probability was consistent for the no-illustration condition regardless of word concreteness (blue line), estimates for the simple-illustration and complex-illustration conditions were positively associated with concreteness (green and red lines, respectively). When concreteness was given a score of 5, estimated performance in both the complex- and simple-illustration conditions exceeded that of the no-illustration condition. Conversely, when concreteness was rated as 1, the no-illustration outperformed the simple-illustration condition, but not the complex-illustration condition. No significant differences between simple and complex illustrations were found at any concreteness level.

Table 7. GLMM Results of Meaning Recall Delayed Posttest with Word Concreteness

Level 1	Level 2	Concreteness	Odds Ratio	95% CI	<i>p</i>
Complex illustration	No-illustration	1	0.38	[0.05, 3.09]	.281
Complex illustration	No-illustration	3	1.16	[0.55, 2.45]	.642
Complex illustration	No-illustration	5	3.53	[1.04, 11.94]	.039
Complex illustration	Simple illustration	1	3.85	[0.43, 34.52]	.281
Complex illustration	Simple illustration	3	1.96	[0.90, 4.27]	.117
Complex illustration	Simple illustration	5	1.00	[0.29, 3.43]	.993
No-illustration	Simple illustration	1	10.17	[1.15, 89.82]	.032
No-illustration	Simple illustration	3	1.69	[0.78, 3.69]	.211
No-illustration	Simple illustration	5	0.28	[0.08, 0.96]	.039

Note. Odds ratios above 1 indicate the superiority of Level 1 over Level 2. *p*-values were adjusted using the Holm method.

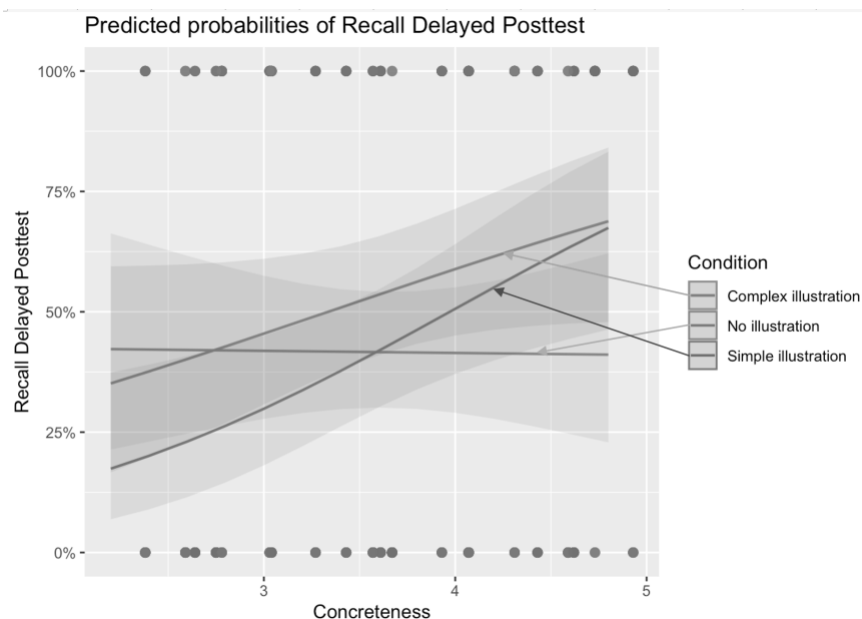


Figure 4. Relationship between Meaning Recall Delayed Posttest and Word Concreteness for Each Learning Condition

Note. No-illustration condition = Target words + L1 translations; Simple-illustration condition = Target words + L1 translations + simple illustrations; Complex-illustration condition = Target words + L1 translations + complex illustrations.

Vocabulary Size. Analyses pertaining to the meaning recall tests indicated that vocabulary size had no significant main effect on either immediate or delayed posttest scores. Additionally, no significant interaction effect was observed between condition and vocabulary size on either immediate or delayed posttest, suggesting that vocabulary size was unrelated to learning gains or the learning condition effects.

RQ3: Which Types of Illustrations do Learners Consider Effective in Vocabulary Learning?

Figure 5 presents the types of illustrations learners consider effective as vocabulary learning aids. Simple illustrations were rated as the most effective by 16 (66.7%) participants, while 11 (45.8%) chose strange illustrations, and only 2 (8.3%) opted for complex illustrations.

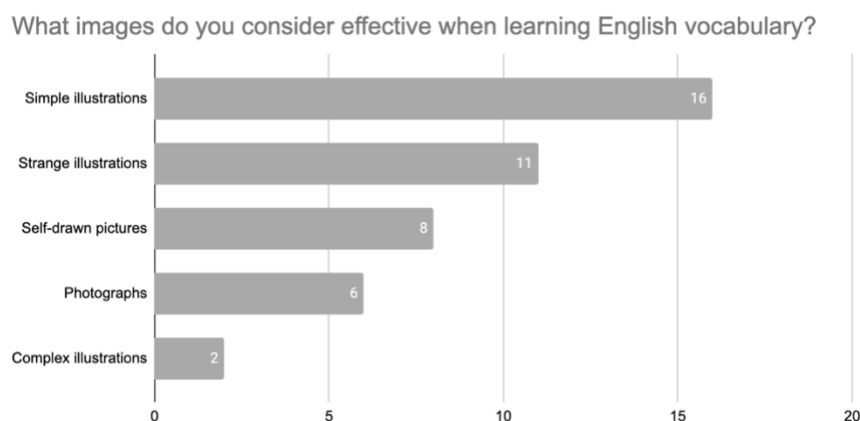


Figure 5. Learners' Perception of Illustration Type Effectiveness for Vocabulary Learning

Note. The results are based on 24 participants' responses. The numbers in the horizontal bars indicate the number of participants that selected the category. Multiple responses were permitted.

Discussion for Experiment 2

RQ1: Do Illustrations Facilitate Vocabulary Learning? If So, Which Type (Complex vs. Simple Illustrations) is Most Effective?

No significant differences were found across different conditions in the immediate posttest, although the comparison of the delayed posttest scores approached statistical significance ($p = .079$) and the mean for meaning recall was consistently the highest for the complex-illustration condition. Especially, on the meaning recall tests, complex illustration led to the greatest learning gains (80.3% and 54.1% for immediate and delayed posttests, respectively), compared to simple illustration (75.0%, 44.6%) and no-illustration (73.1%, 44.0%). Overall, these findings suggest that, using complex illustrations for flashcard-based vocabulary learning potentially leads to greater word learning and retention. However, practically speaking, the effect may not always be clearly impactful as the variance in differences was large across conditions, and the mean score differences remained below 1.

The lack of a clear effect of illustration provision, even with flashcards—where learners presumably paid more attention to illustrations—could be due to the translation-based tests. The transfer appropriate processing theory (Morris et al., 1977) suggests that memory retention increases with higher correspondence between learning and testing tasks. Since both meaning recall and meaning recognition tests required translation for each target word, the no-illustration group may have performed well due to the similarity between learning and testing modalities. While we hypothesized that illustrations could provide additional retrieval routes as learners could recall the illustrations related to target words, which would help them retrieve the corresponding meaning representation, such a positive effect might be negligible.

Another possible explanation relates to cognitive load in the illustration conditions. According to the TOPRA model (Barcroft, 2002; see also Barcroft, 2015), learning multiple aspects at the same time may impose excessive cognitive burden on learners, potentially resulting in failure to acquire any aspect effectively. For example, processing illustrations during learning could have taxed the cognitive resources required for strengthening the connection between a word meaning and form. During the illustration conditions, participants needed to process both translations and illustrations simultaneously, which might have hindered L1 translation acquisition (see Boers et al., 2009 for a similar discussion).

RQ2: To What Extent are Illustration Clarity, Word Concreteness, and Vocabulary Size Related to the Illustration Effect on Vocabulary Learning?

The analysis revealed the interaction effect between illustration conditions and word concreteness. When a word is concrete, clear benefits of providing both simple and complex illustrations were found. However, providing simple illustrations for abstract words hindered learning. During the no-illustration condition, learning performance did not differ regardless of word concreteness. One potential explanation for this finding is that abstract words are difficult to visualize, making it hard to express word meanings with simple illustrations, which thus do not effectively enhance vocabulary learning. Such a clear negative effect was not observed when complex illustrations were used for abstract words. When learning concrete words, learning with illustrations may have supported a clearer understanding of a target word and provided an additional cue to the word's semantic information, fostering additional retrieval routes between words and their meaning representations (Zheng et al., 2016).

Contrary to our hypothesis that abstract words benefit more from picture provision than concrete words, the results indicated that concrete words in fact benefitted more from illustrations. These findings might suggest that illustrations for concrete words may have been more easily retained in memory, thereby increasing the recall of word meanings. An alternative

explanation could be that the illustrations for abstract words tended to express the meanings less clearly, as the additional analysis confirmed that concreteness rating and illustration clarity were significantly and positively correlated both in Experiments 1 and 2. For example the target word “plagiarism” was illustrated with two people; one looking at their own paper, the other looking at the first person’s paper. Such an illustration may have less clearly expressed the word’s meaning, as its clarity was rated low (1.33 on the 5 point-Likert scale). Such a lack of clarity in illustrations for abstract words could have negatively influenced learning. We intended to capture such a negative effect of illustration clarity through our analysis of illustration clarity effects; however, illustration clarity was not significantly related to the learning gains despite its Odds Ratios (ORs) being positive throughout the analyses. This could be attributed to the complex relationship between illustration clarity and concreteness of words, as well as to the limitation of statistical power; we discuss this further in the general discussion and limitation sections.

Given that the effect of concreteness was observed when learning with illustrations, these results showing an unclear influence of illustration clarity may suggest that whether the meaning was clearly expressed by the illustration might not be as influential as concreteness; rather, word concreteness is the determining factor. In other words, even when word meaning is clearly expressed in the corresponding illustration, the meaning of abstract words might still be difficult to recall later on (Ellis & Beaton, 1993). It may thus be useful to further investigate the relationship between illustration features and lexical features in future studies.

Our results showing no clear concreteness effects in the non-illustration condition contrasted with previous studies that found a positive concreteness effect (e.g., de Groot & Keijzer, 2000; Ellis & Beaton, 1993). This discrepancy could be explained by the fact that concreteness rating was based on the L2 target words rather than L1 words or illustrations, potentially not directly reflecting how learners perceive word concreteness. Another possible explanation is that the effects of concreteness are observed in limited contexts. Earlier studies of language learning applications have found no effect of concreteness on vocabulary learning (Hopman et al., 2018; Wild & Kuperman, 2024), implying that the advantage of concrete words on learning might be diminished in retrieval-intensive conditions.

RQ3: What Illustrations do Learners Consider Effective in Vocabulary Learning?

In line with Experiment 1 results, learners expressed their belief that simple illustrations are more effective than the complex ones. Combined with the findings yielded by Experiment 1, these results provide robust evidence that Japanese EFL learners perceive simple illustrations as beneficial for vocabulary learning. While this feedback contrasts with our analysis results indicating a limited impact of illustrations on learning, it is in line with the previous discussion on the motivational benefits of illustration use in learning materials (Katona et al., 2023; Tomlinson, 2008).

General Discussion and Conclusion

For Experiment 1 we used vocabulary lists as the learning aid, while flashcards were adopted for Experiment 2. In Experiment 1, no effects of illustration or other potentially influential factors were observed, with the exception of the positive correlation between vocabulary size and vocabulary learning regardless of illustration provision. In Experiment 2, when solely examining the differences across conditions, no clear advantage of the illustration provision was found. However, the analysis focusing on word concreteness revealed a significant interaction between concreteness and illustration conditions on the meaning recall delayed posttest (but not immediate posttest), indicating that the effectiveness of illustration conditions for vocabulary retention was moderated by word concreteness.

The clearer influence of illustrations observed in Experiment 2 relative to Experiment 1 supports our hypothesis that flashcards facilitated illustration processing by providing additional retrieval opportunities for illustrations relative to word lists (Barcroft, 2007), where illustrations could have been easily ignored. Accordingly, flashcards may have enhanced the retention of the illustrations, which may have served as cues in subsequent tests. It might be hypothesized that flashcards reduce the effect of illustrations compared to list learning, given that participants were exposed to illustrations only when looking at the back of the cards. In contrast, word lists present illustrations to participants for the same duration as target words. Our results, showing the clearer effect of illustrations with flashcards over word lists, highlight that the effect of retrieval opportunity can override the effects of longer exposure to the to-be-retrieved items, corroborating previous studies on the retrieval effect (Barcroft, 2007; Kang et al. 2013).

We attributed the different findings from Experiments 1 and 2 mainly to learning materials (word lists vs. flashcards) and considered that the research design difference (between- vs. within-participants design) may have not been influential to the extent that it overrides the current study's findings. However, because the mean scores across conditions were fairly similar, the influence of illustration in Experiment 2 with increased statistical power might vanish if a between-participants design is adopted. Given this caveat, the impact of illustrations should not be considered as always clearly positive or even pedagogically meaningful, at least in the short term.

Overconfidence may partially explain the less positive effect on abstract words of simple illustrations compared to complex illustrations, and its disadvantageous effects relative to the no-illustration condition (as observed in Experiment 2). In both Experiments 1 and 2, the majority of participants indicated that simple illustrations were effective for vocabulary learning (74.3% and 66.7% for Exp. 1 and Exp. 2, respectively), which was substantially higher than those for complex illustrations (5.1%, 8.3%). Given Carpenter and Olson's (2012) finding that overconfidence can hinder vocabulary learning with illustrations, it is possible that the effect of simple illustrations in this study was compromised because participants believed that the addition of illustrations was facilitative for vocabulary learning. Conversely, when studying with complex illustrations, participants might not have suffered from overconfidence, as they did not consider complex illustrations to be effective. Although we did not directly assess participants' expectations about learning outcomes, expectations might have varied according to this aspect of the illustrations and word concreteness in a complex manner. For example, participants might have considered learning abstract words with simple illustrations easier than with no illustrations or with complex illustrations. It would be interesting to more closely investigate potential effects of perceived effectiveness in future studies.

It is also important to consider potential reasons for variables that are not significantly associated with learning. The lack of any clear effect of illustration clarity could partially be due to insufficient statistical power. The ORs exceeded 1 by a greater margin for immediate and delayed posttests in both Experiments 1 and 2. This suggests the modest but positive association between the clarity of illustration and learning gains. The lack of statistical power may also explain the non-significant effect of vocabulary size on learning in Experiment 2. By adopting a within-participant design with a total of 24 participants, the vocabulary size of each participant was limited for detecting its stable and clear influence. Thus, it is important for future studies to include a larger sample size to capture the modest but clear effects of illustration clarity and to analyze the potential interaction effect between vocabulary size and illustration conditions.

Limitations and Suggestions for Future Research

While we focused on single nouns as target vocabulary, expanding this line of research to other parts of speech, multi-word units, and idioms may help to evaluate the generalizability of the current findings (see Boers et al., 2009). Similarly, learning was measured only at the meaning recall and recognition levels. As the TOPRA model predicts, even the negative effects of illustration learning could have been observed when measuring learning at the form recall and recognition levels (Barcroft, 2015). Thus, the findings reported here need to be interpreted with caution, as they might only pertain to the learning of the meaning aspects of vocabulary knowledge. The generalizability of our results could also be limited due to the unique characteristics of our participants. The experiments were conducted at a nationally ranked university, where the entrance exam is highly competitive and requires intensive rote memorization skills. This suggests that high rote memorization skills may have led to high scores even on the delayed posttests, and points to the possibility that learners varying in rote memorization skills might exhibit different effects of illustrations on learning. Furthermore, as most young learners are still developing their L1 knowledge, they might benefit more clearly from the illustrations in addition to their purely motivational benefits (Ramezani et al., 2021).

As for word-related factors, we focused on concreteness. However, word concreteness may relate to other potentially influential factors such as pronounceability, imageability, corpus frequency, and word length (see Lindstromberg & Boers, 2023, for a similar discussion of the complex relationship between word-related factors). It would be useful for future studies to explore the different word-related factors that potentially play a role in deliberate vocabulary learning with illustrations. Additionally, we adopted the concreteness ratings of English target words provided by Brysbaert et al. (2014). However, given that the participants had little prior knowledge of the target words, ratings based on their L1 translations (and illustrations) may have been more appropriate. Further investigation with L1-based ratings would be important for a more accurate estimation of the concreteness effect.

We operationalized complexity as a dichotomous variable (simple vs. complex) denoting different amounts of information related to illustrations. While we believe this operationalization provides useful evidence on this topic and has greater ecological validity than comparing pictures with slight differences related to a specific variable, future studies should explore which specific types of complexity (e.g., contextual information, degree of realism, detail) are closely related to learning.

As discussed previously, the variables found to be non-significant could be attributed to a lack of power to detect their effects. We conducted simulation-based posterior power analyses with the *simr* package to obtain an idea of an adequate sample size (power > .8 with alpha = .05 for the effect sizes observed in the current study) for future replication studies to detect the significant mean differences between illustration conditions, and the effect of illustration clarity on the immediate meaning-recall posttest. The results suggest that to detect the main effects of illustration conditions with sufficient statistical power, the fitted models of Experiment 1 required a total of 1222 participants, and the models of Experiment 2 needed 64 participants.

Regarding illustration clarity, some of the illustrations did not clearly express the meaning of their target words (i.e., plagiarism). Although we hoped that our analysis could detect a clear relationship between the clarity of illustration and learning gains, the results failed to find any meaningful relationship. The results of power analysis showed that to detect the main and interaction effects of illustration clarity on the immediate posttest, the fitted models of Experiment 1 required a total of 190 participants for the main effect and 183 participants for the interaction effect. Similarly, the models of Experiment 2 needed 132 participants for the

main effect and 142 participants for the interaction effect (see OSF for the detailed results). Additionally, future studies may also benefit from simply replacing unclear illustrations with clearer ones to maximize the potential benefits of illustrations.

Pedagogical Implications

Based on the analysis of our limited data, for university-level EFL learners learning nouns, the presence or absence of illustrations may not lead to significant differences in short-term learning outcomes, especially with list learning. However, word concreteness may be considered as the one moderating factor altering the effect of illustrations. For concrete words, the use of illustrations is recommended, whereas for abstract words, illustrations may not be necessary, and L1 translations alone may suffice. Simple illustrations should be avoided for abstract words. As learners generally have positive attitudes toward illustration use for vocabulary learning, complex illustrations for concrete words may justify the cost.

Acknowledgments

This study is based on the first author's Master's thesis submitted to the University of Tsukuba. We are grateful to Rie Koizumi, Yuko Hijikata, and Joe Barcroft for their valuable comments on the study. We would like to thank the TESL-EJ editors and anonymous reviewers for their helpful feedback on earlier versions of this manuscript.

About the Authors

Satoshi Ide earned his Master's degree in Literature from the University of Tsukuba, Japan, where he specialized in English Language Education. His research interests include deliberate vocabulary learning and the development of educational materials. He has presented his research at EuroSLA, AsiaTEFL, and the Hiroshima Lexical Research Forum. ORCID ID: 0009-0009-9643-3727

Akifumi Yanagisawa is Assistant Professor at the Institute of Humanities and Social Sciences at the University of Tsukuba, Japan. His research focuses on second language vocabulary acquisition, and he is particularly interested in cognitive factors that influence vocabulary learning. His work has appeared in journals such as *Language Learning*, *Studies in Second Language Acquisition*, and *The Modern Language Journal*. ORCID ID: 0000-0002-7769-5758

To Cite this Article

Ide, S. & Yanagisawa, A. (2026). What types of illustrations are effective in intentional vocabulary learning? Simple or complex? Clear or unclear? *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 30(1). <https://doi.org/10.55593/ej.30117a5>

References

- Alsalihi, H. D. (2020). Posters in vocabulary learning. *Arab World English Journal: Special Issue on the English Language in Iraqi Context*, 18–31. <https://doi.org/10.24093/awej/elt2.2>
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). International Reading Association.
- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323–363. <https://doi.org/10.1111/0023-8333.00186>

- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, 57(1), 35–56. <https://doi.org/10.1111/j.1467-9922.2007.00398.x>
- Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins. <https://doi.org/10.1075/llt.43>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bates, J., & Son, J.-B. (2020). English vocabulary learning with simplified pictures. *TESL-EJ*, 24(3), 1–20. <https://tesl-ej.org/wordpress/issues/volume24/ej95/ej95a12>
- Boers, F., Piquer Píriz, A. M., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of idioms? *Language Teaching Research*, 13(4), 367–382. <https://doi.org/10.1177/1362168809341505>
- Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does adding pictures to glosses enhance vocabulary uptake from reading? *System*, 66, 113–129. <https://doi.org/10.1016/j.system.2017.03.017>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 92–101. <https://doi.org/10.1037/a0024828>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- de Groot, A., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, 50(1), 1–56.
- Dziemianko, A. (2022). The usefulness of graphic illustrations in online dictionaries. *ReCALL*, 34(2), 218–234. <https://doi.org/10.1017/S0958344021000264>
- Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43(4), 559–617. <https://doi.org/10.1111/j.1467-1770.1993.tb00627.x>
- Goodman, M. S., Lopez, A., Murillo, A. L., & Pierce, K. A. (2025). A comparison of methods for coding race in linear and logistic regression models. *Annals of Epidemiology*, 112, 15–22. <https://doi.org/10.1016/j.annepidem.2025.10.005>
- Hamada, A., Iso, T., Kojima, M., Aizawa, K., Hoshino, Y., Sato, K., Sato, R., Chujo, J., & Yamauchi, Y. (2021). Development of a vocabulary size test for Japanese EFL learners using the New JACET List of 8,000 basic words. *JACET Journal*, 65, 23–45. https://doi.org/10.32234/jacetjournal.65.0_23

- Hopman, E., Thompson, B., Austerweil, J. L., & Lupyan, G. (2018). Predictors of L2 word learning accuracy: A big data investigation. *Proceedings of the 40th Annual Conference of the Cognitive Science Society (CogSci 2018)*. <https://par.nsf.gov/biblio/10074358-predictors-l2-word-learning-accuracy-big-data-investigation>
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207–223.
- Inoue, N., & Akano, I. (2003). *The Wisdom English-Japanese dictionary*. Sanseido (Wisdom).
- Ishikawa, S. (2018). Development of the advanced English academic vocabulary list “BABILON 2000”: A new attempt at EGAP vocabulary selection based on six principles. In Y. Ishikawa (Ed.), *Horizons of ESP Vocabulary Studies* (pp. 2–20). Taishukan.
- Jeon, E. H., & In'nami, Y. (Eds.). (2022). *Understanding L2 Proficiency: Theoretical and meta-analytic investigations* (Vol. 13). John Benjamins Publishing Company. <https://doi.org/10.1075/bpa.13>
- Kang, S. H. K., Gollan, T. H., & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review, 20*(6), 1259–1265. <https://doi.org/10.3758/s13423-013-0450-z>
- Kaplan-Rakowski, R. (2019). The effect of stereoscopic three-dimensional images on vocabulary learning. *Contemporary Educational Technology, 10*(4), 324–337. <https://doi.org/10.30935/cet.634172>
- Kasahara, K., & Yanagisawa, A. (2024). Learning new verbs with known cue words: The relative effects of noun and adverb cues. *Language Teaching Research, 28*(1), 138–155. <https://doi.org/10.1177/13621688211001613>
- Katona, B., Venkataragavan, J., Nina, E., Ulrika, B., & Björn, O. (2023). Use of visual learning media to increase student learning motivation. *World Psychology, 1*(3), 161–176. <https://doi.org/10.55849/wp.v1i3.381>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lindstromberg, S., & Boers, F. (2023). Word classes in second language acquisition. In E. Van Lier (Ed.), *The Oxford Handbook of Word Classes* (1st ed., pp. 876–886). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198852889.013.37>
- Mahdi, H. S., & Gubeily, M. A. I. (2018). The effect of using bizarre images as mnemonics to enhance vocabulary learning. *Journal of Social Studies, 24*(1), 113–135. <https://doi.org/10.20428/jss.v24i1.1310>
- Maley, A. (2011). Extensive reading: Maid in waiting. In B. Tomlinson (Ed.), *English Language Learning Materials: A Critical Review* (pp. 133–156). Continuum.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519–533. [https://doi.org/10.1016/S0022-5371\(77\)80016-9](https://doi.org/10.1016/S0022-5371(77)80016-9)

- Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL*, 20(1), 3–20. <https://doi.org/10.1017/S0958344008000219>
- Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, 41(2), 287–311. <https://doi.org/10.1017/S0272263118000219>
- Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
- OpenAI. (2024). *ChatGPT* (December version) [Large language model]. <https://chat.openai.com/>
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 241–263. <https://doi.org/10.1037/h0027272>
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11(4), 137–138. <https://doi.org/10.3758/BF03331011>
- Paivio, A., & Yarmey, A. D. (1966). Pictures versus words as stimuli and responses in paired-associate learning. *Psychonomic Science*, 5(6), 235–236. <https://doi.org/10.3758/BF03328369>
- Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, 55(1), 105–133. <https://doi.org/10.1002/tesq.579>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 156–163. [https://doi.org/10.1016/S0022-5371\(67\)80067-7](https://doi.org/10.1016/S0022-5371(67)80067-7)
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Skulmowski, A., & Rey, G. D. (2018). Realistic details in visualizations require color cues to foster retention. *Computers & Education*, 122, 23–31. <https://doi.org/10.1016/j.compedu.2018.03.012>
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215. <https://doi.org/10.1037/0278-7393.6.2.174>
- Tomlinson, B. (Ed.). (2008). *English language learning materials: A critical review*. Continuum.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 1–17.

- Webb, S. (2024). Methodological features of studies of incidental vocabulary learning. In M. Feng Teng & B. L. Reynolds, *Researching incidental vocabulary learning in a second language* (pp. 53–68). Routledge. <https://doi.org/10.4324/9781003270782-6>
- Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651–675. <https://doi.org/10.1017/S0272263114000606>
- 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>
- Wild, H. A., & Kuperman, V. (2024). Word learning in the wild: App-based evidence for valence and concreteness effects. *Applied Psycholinguistics*, 45(5), 786–810. <https://doi.org/10.1017/S0142716424000304>
- Wright, A. (1989). *Pictures for language learning*. Cambridge University Press.
- Yowaboot, C., & Sukying, A. (2022). Using digital flashcards to enhance Thai EFL primary school students' vocabulary knowledge. *English Language Teaching*, 15(7), 61–74. <https://doi.org/10.5539/elt.v15n7p61>
- Zheng, J., Zhang, W., Li, T., Liu, Z., & Luo, L. (2016). Practicing more retrieval routes leads to greater memory retention. *Acta Psychologica*, 169, 109–118. <https://doi.org/10.1016/j.actpsy.2016.05.014>

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.