Studies in Second Language Acquisition (2025), **47**, 157–180
doi:10.1017/S0272263124000718

RESEARCH ARTICLE

Contextual learning and retention of phrasal verbs

The effects of definition placement and typographic enhancement

Mojtaba Tadayonifar, Irina Elgort and Anna Siyanova-Chanturia

School of Linguistics and Applied Language Studies, Te Herenga Waka–Victoria University of Wellington, Wellington, New Zealand
Corresponding author: Mojtaba Tadayonifar; Email: mtadayon.253@gmail.com

(Received 16 February 2024; Revised 17 October 2024; Accepted 25 November 2024)

Abstract

A common way of acquiring multiword expressions is through language input, such as during reading and listening. However, this type of learning is slow. Identifying approaches that optimize learning from input, therefore, is an important language-learning endeavor. In the present study, 85 learners of English as a foreign language read short texts with 42 figurative English phrasal verbs, repeated three times. In a counterbalanced design, we manipulated access to definitions (before text, after text, no definition) and typographic enhancement (with bolding, without bolding). The learning was measured by immediate and delayed gap-fill and meaning generation posttests. All posttests showed that learning with definitions was better than without, and that access to definitions *after* reading was more beneficial than *before* reading. Typographic enhancement effectively promoted contextual learning of phrasal verbs and increased the learning advantage associated with presenting definitions after reading.

Keywords: contextual learning; definitions; phrasal verbs; retention; typographic enhancement

Introduction

Learning vocabulary in a second language (L2) can take place out of context, for example, using vocabulary lists and flashcards; or it can happen in context, for instance, during reading and listening, with or without additional vocabulary learning support. Research has shown that multiword expressions (MWEs[1]) can be learned incidentally from reading and listening (e.g., Toomer & Elgort, 2019; Webb, Newton, & Chang, 2013). However, L2 vocabulary gains in contextual learning without support tend to be

[1]Multiword expressions (MWEs) have been defined as conventional strings of language above the word level (Siyanova-Chanturia & Pellicer-Sánchez, 2019).

small with novel vocabulary taking a long time to be acquired (e.g., Pavia, Webb, & Faez, 2019; Webb, Uchihara, & Yanagisawa, 2023). One way to facilitate contextual learning is through access to definitions of novel vocabulary (AbuSeileek, 2011; Bolger, Balass, Landen, & Perfetti, 2008). However, whether it is more beneficial to access definitions before or after reading is only now starting to be addressed (Elgort, Beliaeva, & Boers, 2020). Recent studies have shown that previewing and preteaching novel words before encountering them in reading changes how readers engage with these words during reading (Elgort, van de Wetering, Arrow, & Beyersmann, 2023; Pellicer-Sánchez, Conklin, & Vilkaitė-Lozdienė, 2021), which, in turn, may affect their learning. However, when novel vocabulary is not previewed, readers' incorrect contextual inferences may result in encoding erroneous form-meaning mappings, which may hinder future learning (Yu & Boers, 2023). Therefore, the question of whether contextual inferences should be preceded or followed by definitions is a matter that requires further empirical evidence.

Another technique that is known to affect attention to novel vocabulary in reading is the use of typographic enhancement, such as bolding or underlining. Typographic enhancement has been shown to facilitate contextual learning of MWEs (El-Dakhs et al., 2021; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019), especially less perceptually salient types, such as grammatical collocations (Toomer, Elgort, & Coxhead, 2024), presumably because it draws learners' attention to the whole MWE during reading, which is otherwise easy to miss. Thus, typographic enhancement may boost the positive effect of definitions in contextual learning of less perceptually salient types of MWEs, such as figurative phrasal verbs (e.g., "*hold up*," "*figure out*").

The present study investigates whether definition placement and typographic enhancement affect contextual learning of figurative phrasal verbs (PVs). PVs are one of the most difficult types of MWEs to learn contextually because they consist of a lexical verb and an adverbial particle that is not salient in the input (Gardner & Davies, 2007). Learning figurative PVs from reading is particularly challenging (El-Dakhs et al., 2021) because their meaning senses cannot be easily inferred from the meanings of their parts or context.

Background

Previous studies that have investigated incidental learning of MWEs from input have yielded conflicting results. Some studies have shown that repeated exposure to MWEs can significantly enhance learners' acquisition of these items (e.g., Pellicer-Sánchez, 2017; Puimège & Peters, 2019; Sonbul & Schmitt, 2013; Vu & Peters, 2022; Webb et al., 2013). However, incidental learning of MWEs from input was not significant in Szudarski (2012) and Szudarski and Carter (2016). Reviewing 24 primary studies, Webb et al. (2023) found similar results for gains in L2 vocabulary learning from reading (between 15–17 percent), listening (between 13–15 percent), and reading while listening (between 13–17 percent) conditions. These findings suggest that learning L2 MWEs from input only may be difficult. This is partly because reading in a second language is a daunting task for L2 learners due to the high proportion of unknown words in nonsimplified text (Zhang & Ma, 2021). Another reason is that lower-proficiency L2 learners take longer than advanced L2 learners to establish lexical representations from reading or listening only (Elgort & Warren, 2014). Therefore, instructional and learning support may be needed to facilitate L2 vocabulary learning from input, especially for lower-proficiency learners. One such type of support is the provision of definitions.

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

The effects of definitions on contextual vocabulary learning

In contextual vocabulary learning from reading, providing definitions facilitates the abstraction of word meanings from specific contexts, according to the instance-based memory model (Reichle & Perfetti, 2003). Thus, every time learners encounter a novel word in a text, an episodic memory trace for that word plus the context in which it was used is established. After multiple encounters with the novel word in different contexts, the overlapping features of its meaning are consolidated and the nonoverlapping (e.g., erroneous) aspects are discarded, resulting in the establishment of a core meaning of that word. In other words, providing definitions facilitates the process of establishing a core meaning for the target words and of abstraction of that meaning from individual contexts in which the word had been previously experienced. Because dictionary-type definitions contain core semantic features of a word, access to definitions in contextual learning can be described as a super learning instance (Bolger et al., 2008).

Previous L2 vocabulary learning studies targeting single words have obtained evidence in favor of providing definitions compared with reading-only conditions (e.g., AbuSeileek, 2011; Elgort et al. 2020; Hulstijn, Hollander, & Greidanus, 1996; Zhang & Ma, 2021). For example, Elgort et al. (2020) showed that when definitions were not presented during the learning procedure, incorrect contextual meaning inferences negatively impacted the development of declarative word knowledge. Therefore, Elgort et al (2020) suggested that a brief familiarization with the target words through definitions might be beneficial for contextual vocabulary learning.

Although the benefits of providing definitions in contextual vocabulary learning are not controversial, the issue of *when* to provide them has been less studied. Providing definitions after reading encourages learners to infer the meaning of novel words during reading. Inferencing enhances learning as it requires a certain degree of cognitive effort (Yu & Boers, 2023), which leads to improvements in long-term retention (Bjork & Bjork, 2014). In learning and memory research, lexical inferencing has been considered an example of the "generation effect," defined as a phenomenon in which memory for generated information is stronger than for the information that is simply read (Bertsch, Pesta, Wiscott, & McDaniel, 2007). For example, if the goal is to memorize an antonym of the word *hot*, based on the generation effect, it is more effective to ask learners to generate an antonym for it (*cold*), compared with simply providing both words and asking learners to read them (see Bertsch et al., 2007 for a meta-analysis).

Inference-making involves guessing the meanings of the target items, which may serve as semantic elaboration, a process shown to lead to durable memory traces (Craik & Tulving, 1975). To measure the durability of memory traces, Craik and Tulving (1975) used structural, phonetic, and semantic questions. The results of an uninformed recall test showed that recognition accuracy was higher after semantic questions. They attributed this to the higher degree of stimulus elaboration after semantic questions than after structural and phonetic questions.

L2 vocabulary research has also found that inferring novel word meanings during reading, followed by definitions, benefits contextual word learning (e.g., Elgort et al., 2020; Huang & Lin, 2014). Huang and Lin (2014) exposed Chinese EFL learners to a text containing a set of novel L2 words repeated three times. In one condition, L1 translations were provided for all occurrences of the L2 words, while in the other condition, translations were given only after the second occurrence, requiring learners to infer meanings initially. The posttest results indicated that learners in the inference condition demonstrated superior recall of meanings. Elgort et al. (2020) tested the

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

effects of definition placement on the contextual learning of single words. They instructed 55 L1 English speakers and 52 Chinese ESL learners to read short texts that included three instances of 90 novel vocabulary items and infer their meanings from context. In this study, definitions of the target items were given before reading the texts, after reading the texts, or were not given. The results showed that presenting definitions after the texts resulted in superior vocabulary learning outcomes in the posttests for L1 and L2 learners compared to the other conditions. They attributed the advantage of presenting definitions after reading to the learners' mental effort needed to infer the meanings of the critical items, which might have led to deeper encoding of the items. In a follow-up eye-tracking study, Elgort et al. (2023) found that learners spent more time reading the critical items when definitions were presented after rather than before reading.

Presenting definitions after reading, however, may sometimes lead to erroneous meaning inferences. There is a concern that incorrect meaning inferences may interfere with later meaning recall and hence slow down the acquisition of L2 words. Several L2 studies on novel idioms have found that incorrect inferences were retained even when corrective feedback was provided and interfered with the acquisition of correct meanings of idioms (Wang, Boers, & Warren, 2022; Yu & Boers, 2023). Yu and Boers (2023) compared providing definitions for idiomatic expressions before (meaning given) or after (inferencing) the text. The inferencing condition was further divided into two subcategories, one aimed at increasing the likelihood of correct inferences by providing literal underpinnings of idioms. In the other condition, participants were shown an example of the idiom in context and then prompted to offer their understanding of its meaning. The results of an unannounced recall test after a week revealed no significant difference between providing definitions before the text and the inferencing-first condition, where chances of making incorrect inferences were high. However, in the second inferencing condition, where the likelihood of making incorrect inferences was low, participants performed significantly better than those in the meaning-given condition. The findings of Yu and Boers (2023) are aligned with Elgort et al. (2020), who also found better learning outcomes when contextual inferences were correct compared to incorrect inferences. Therefore, further studies are needed to clarify if inferring meanings from context is preferable to familiarization with definitions prior to contextual learning. In the case of contextual learning of novel figurative MWEs (such as idioms and figurative PVs), one reason for incorrect inferences and poor form-meaning mappings could be learners' failure to notice novel expressions, as a whole, and attempting to fit the meanings of their component words into context. The use of learning interventions that affect attention to and memory of whole MWEs in reading, such as typographic enhancement, may provide a boost to improve the accuracy of contextual inferences.

Typographic enhancement and contextual vocabulary learning

Typographic enhancement involves highlighting target items (words or phrases) by making use of underlining, bold typeface, italics, or uppercase (Campillo, 2015), to render them more noticeable in the input than they would be without enhancement. The noticing hypothesis argues that what learners deliberately attend to or notice unintentionally in the input is what becomes intake (Schmidt, 2001). To explore the learning and processing of textually enhanced MWEs, Choi (2017) tested the effects of textual enhancement on the learning of L2 collocations. The study further aimed to

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

check whether enhancing MWEs affects recalling unenhanced text. To do so, 38 Korean EFL learners were divided into two groups. Then, two versions (enhanced and unenhanced) of 10 texts with 14 target collocations were developed. One group read the enhanced texts, and the other group read the unenhanced texts while their eye movements were recorded. On the postreading collocation test, the participants in the enhanced condition outperformed the unenhanced group. However, Choi (2017) found a trade-off between learning collocations in the enhanced condition and recalling the unenhanced text. The participants reading typographically enhanced texts recalled significantly less unenhanced text than the group who read the unenhanced texts. Eye fixation data revealed that the participants in the enhanced condition spent substantially longer time processing unfamiliar collocations.

More recently, Puimège, Montero Perez, and Peters (2023) conducted an eye-tracking study to explore the impact of typographic enhancement on contextual learning and processing of L2 MWEs. The participants (61 Dutch-speaking students learning English as their L2) were split into experimental and control groups. The experimental group read 10 English texts containing 24 transparent modifier-noun (e.g., sensitive cells) collocations in which 12 target collocations were enhanced and the other 12 were not. These collocations were repeated eight times. The control group read a version of the same texts that did not contain the target collocations. They showed that typographic enhancement was effective in drawing learners' attention to the target items in the first exposure, as the enhanced collocations obtained significantly longer reading times. The eye movement results revealed that even though typographic enhancement had an initial influence on learners' perceptual processing of the critical items, it failed to attract attention to the target collocations, as the positive effects did not carry over to later, unenhanced exposures. They further revealed that employing typographic enhancement did not prompt the degree of attention required to result in a durable memory trace, as the majority of the participants made no attempts to memorize the target-enhanced collocations.

Although typographic enhancement has been found to influence the early stages of encoding information into memory, it may not lead to robust learning and durable learning outcomes (e.g., Northbrook, Allen, & Conklin, 2022; Szudarski & Carter, 2016), especially for less transparent vocabulary items (e.g., Campillo, 2015; El-Dakhs et al., 2021). For example, Campillo (2015) tested the effectiveness of typographic enhancement for the form recognition and comprehension of transparent (e.g., *to be a bag of bones*) and nontransparent (e.g., *wet blanket*) English idioms. To this end, participants were exposed to enhanced and unenhanced L2 idioms in short texts. The results showed that typographic enhancement did not have a positive effect on the recognition of opaque idioms. El-Dakhs et al. (2021) also investigated the efficacy of enhanced conditions on the incidental learning of transparent (e.g., *bring in*) and opaque (e.g., *take in* meaning *deceive*) PVs. The participants were divided into three groups. The incidental group received a text with the target PVs presented in normal font. The enhanced group read the same text, while the target PVs were underlined and bolded; the control group followed their usual learning condition with no experimental treatment. The results showed that the participants in the enhanced exposure condition outperformed the incidental condition. They also found that input enhancement that directed learners' attention to the PVs had a positive effect on the learning of transparent but figurative PVs, likely because the meanings of figurative PVs were more difficult to infer from context. Thus, while noticing might be a basic requirement for learning, it does not necessarily guarantee the acquisition of the noticed MWEs (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006). These results suggest that, for

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

figurative PVs, the provision of definitions, alongside typographic enhancement, may speed up contextual learning.

Combining definitions and typographic enhancement in contextual vocabulary learning

Research on the combined effects of typographic enhancement and provision of definitions on vocabulary learning, especially for MWEs, is limited. In Peters (2012), the experimental group read a text (with L1 definitions in the margins) and then wrote down MWEs. The control group was instructed to read the same text and write down unfamiliar vocabulary, without any reference to MWEs. In both conditions, half of the target items were typographically enhanced. The results of form recall posttests showed that typographic enhancement was more effective in facilitating participants' noticing and learning of unfamiliar MWEs than explicitly instructing them to focus on these expressions. It was further found that definitions drew learners' attention to the target items. Qualitative data obtained from a questionnaire showed that definitions also helped learners to carry out their vocabulary task sheets and to prepare them for the upcoming posttest.

In a similar study, albeit with a more extensive range of experimental conditions, Toomer and Elgort (2019) exposed L2 English learners to a text that contained low-frequency medical collocations (e.g., *cloud baby*) and their L2 explanations under reading only, typographic enhancement and typographic enhancement plus definition conditions (definitions were presented in the margins and were different from the contextual explanations). They found that adding definitions to the typographic enhancement did not create a learning advantage. This was contrary to their hypothesis that providing definitions to the enhanced items leads to superior learning. Toomer and Elgort (2019) attributed this to the interruption that occurred when learners took their eyes away from the text to read the definitions. Boers (2020) argued that providing definitions in an early encounter with a text might affect learners' engagement with the target items on subsequent reencounters. However, the impact of definitions was not deliberately manipulated in previous studies.

To further investigate the effects of definition placement on the contextual learning and online processing of words, Elgort et al. (2020) and Elgort et al. (2023) conducted two single-word learning studies. In Elgort et al. (2020), the target items were presented in brackets to emulate a condition in which the novel words are noticed. The results showed that presenting definitions after the texts resulted in better word knowledge compared to presenting definitions before the texts for both L1 speakers and L2 learners. To seek evidence on whether previewing novel words before reading may affect how attention is allocated to these words during reading, Elgort et al. (2023) conducted an eye movement study. L1 and L2 speakers of English read passages that contained 60 novel words. In this study, the target items were not typographically enhanced or explicitly identified in the text in any way. The authors found shorter reading times and higher skipping rates on the novel words when definitions were previewed, relative to viewing definitions after reading. This confirmed that learners paid less attention to the previewed words during reading, relative to the condition when novel words were first encountered in reading.

In Elgort et al. (2023), however, the effect of definition placement was only observed on the gap-fill posttest which used supportive contexts (and only for L1 participants), but not on a meaning generation test when participants had to recall meanings of the

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

target items without contextual support. Note that, in Elgort et al. (2023), the target items were not identified in the passages in any way. Thus, we conjecture that explicitly identifying target vocabulary in reading may have boosted the positive effect of contextual inferencing in the condition where definitions were presented after reading. In other words, employing a method that directs learners' attention towards the target vocabulary items during reading (e.g., bolding) may make novel items more salient in the text, facilitating contextual inferencing when novel items appear first in reading. The present study was designed to orthogonally manipulate definition placement and typographic enhancement to test this hypothesis and to extend previous findings to contextual learning of figurative PVs.

The present study

The present study manipulated the provision of definitions, their placement (before reading, after reading), and the use of typographic enhancement (with bolding, without bolding) to establish optimal conditions for contextual learning of L2 figurative PVs while participants read short texts. The immediate and delayed posttests of form and meaning recall were used to measure initial learning and retention of the PVs.

The following research questions (RQs) were posited:

RQ1. Is contextual learning and retention of PVs from reading affected by the provision of definitions?
RQ2. Is contextual learning and retention of PVs from reading affected by definition placement?
RQ3. Is contextual learning and retention of PVs from reading affected by typographic enhancement?
RQ4. Does the effect of typographic enhancement modulate the effect of definition placement?

Methodology

Participants

Eighty-five high school students between 16–18 years of age participated in this study. These participants were in five intact classes. They were Persian L1 speakers learning English as a foreign language. They had studied English for at least three years at junior high school. They had English lessons twice a week. All classes were using the same textbooks. Their English proficiency was determined using the Preliminary English Test (PET). According to the Cambridge scoring (based on the online calculator using https://cambridgescore.com/pet), the test ranking is as follows: 1–12, Elementary (A1); 13–22, Preintermediate (A2); 23–28, Intermediate (B1); and 29–32, Upper Intermediate (B2). According to the data (M = 16.31, SD = 3.39), 84% of participants were considered preintermediate English learners, 12% as elementary, and 4% as intermediate.

Materials

Target phrasal verbs

A key criterion in selecting the target items was that individual words within PVs should be known to the study participants. As the participants were low-proficiency

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

English learners, only PVs with individual words within the first 2000 most frequent words of English (based on the Corpus of Contemporary American English [COCA]) were included in the study. In total, 120 PVs that had an MI score[2] (as a measure of association strength) of 3 and above were chosen from corpus-based lists of PVs (e.g., Gardner & Davies, 2007), and textbooks containing PVs (e.g., McCarthy & O'Dell, 2004). These items underwent three norming procedures before the final items were selected. In the first norming procedure and to select figurative PVs, 33 English L1 speakers rated (in two surveys) the extent to which the meaning of the phrase (as used in the given sentence) was the same as the meaning of its components put together, on a seven-point Likert scale (1 = fully literal and 7 = fully figurative). The items with a mean figurative score of 4 and above were selected (n = 72).

In the second norming procedure, 24 L2 learners (similar in characteristics to the participants in the main study) were instructed to explain the meaning of 72 PVs, as well as the meaning of the first word of the PVs, either in Persian (L1) or in English (L2). The items, for which 80% of the participants did not know the figurative meanings but knew the meaning of their first word (verb), were selected (n = 62).

In the third norming procedure, four teachers, whose students were participating in the study, were asked to indicate whether their students were likely to know the meaning of the target PVs. The items (n = 42), which three out of four teachers rated as likely unknown, were selected for the study. As a result of the above norming procedures, 42 PVs were selected (e.g., *chip in*) for the experiment. (Here is the link in the Open Science Framework (OSF) to access the materials: [https://osf.io/yzwdh/?view_only=8c75029506034b44bbf9349bc056b0b9]. This link provides access to all the relevant materials, data, and data analysis used in the study, ensuring transparency and facilitating further research and replication.)

The texts

Forty-two texts were developed while controlling for word frequency, length of words, and readability. Each PV was repeated three times in the same text. AntWordProfiler (Anthony, 2022) was used to check lexical frequency. To allow for the inclusion of the PVs as novel items while reaching 98% lexical coverage (Nation, 2013), all of the other words used in the texts were selected from the first 2000 words, including proper nouns and transparent compound lists. The participants showed mastery of the first 2000 words based on the results of the Updated Vocabulary Levels Test (see below). The developed texts were between 77 to 92 words in length (M = 84.04, SD = 4.2). The Flesch–Kincaid grade level (Kincaid, Fishburne, Rogers, & Chissom, 1975) was used to calculate the readability scores (M = 5.48, SD = 1.11) using an online instrument (https://charactercalculator.com/flesch-reading-ease/, 2023). It shows the number of years of education required to understand a text for English L1 speakers. Based on the readability scores, the texts were found suitable for 5th to 7th graders.

No object was used between the verb and the particle (e.g., *back up ideas*). This is because research has shown that L2 learners may process MWEs differently when they are adjacent (e.g., *hand over* the responsibility) than when they are nonadjacent (e.g., *hand* the responsibility *over*) (Vilkaite & Schmitt, 2017). To increase the likelihood that the initial stages of learning have occurred, the target items appeared three times, in the same meaning sense, in the texts (Elgort et al., 2020).

[2]This threshold is arbitrary and has been criticized by Eguchi & Kyle (2023).

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

As the present study intended to test the effect of providing definitions on the contextual learning of figurative PVs, the texts did not strongly constrain the meaning of the target items (e.g., care was taken not to use synonyms or other words that could reveal the figurative meanings of the PVs from reading only). Ten L1 English speakers reviewed the texts, highlighting any words that could give away the meanings of the target PVs for learners to infer the meanings; the texts were further revised based on the feedback received. The words that could potentially give away the meaning of the target items were either deleted or replaced with other, less revealing words.

Experimental design

The current study involved a 2 × 3 within-participant design. The independent variables were typographic enhancement with two levels (enhanced, unenhanced), and definition placement with three levels (before text, after text, no definition). The dependent variables were the scores from immediate and delayed form and meaning recall posttests. Table 1 shows the learning design for this study.

Pretests

The Preliminary English Test (PET)

The Preliminary English Test (PET) was used as the proficiency test. PET is an English language examination supported by Cambridge Assessment English (https://www.cambridgeenglish.org/exams-and-tests/preliminary/). The test has reading, writing, listening, and speaking sections. The reading section which consisted of six parts and 32 questions was administered.

The Updated Vocabulary Levels Test (UVLT)

The Updated Vocabulary Levels Test (UVLT) (Webb, Sasao, & Ballance, 2017) was administered to estimate participants' L2 vocabulary knowledge. There are five levels in UVLT, and each level consists of 30 questions. To complete this test, the participants need to match each definition to the word it defines. All the five levels were administered. The minimum score required to demonstrate mastery of a vocabulary level (for VLT) appears to have been arbitrary (Xing & Fulcher, 2007). While Webb et al. (2017) recommend a cutting point of 29/30 at the 1000, 2000, and 3000 levels (for UVLT), Schmitt, Schmitt, and Clapham (2001) recommend 80% (24/30) as a mastery threshold (for VLT). Thus, the scores of 26 and 25 on levels 1 and 2 respectively were considered to be sufficient to confirm participants' mastery of these levels. The participants showed mastery of the first (M = 26.3, SD = 1.5) and the second word levels (M = 25.4, SD = 1.2).

Table 1. Experimental design

Learning conditions	Definition placement	Typographic enhancement
1	Before text	Yes
2	Before text	No
3	After text	Yes
4	After text	No
5	None	Yes
6	None	No

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

Posttests

Gap-fill test

A cued recall test based on Garnier and Schmitt (2016) was used to measure the participants' knowledge of the form of the target PVs. It was a pen-and-paper form recall test in the form of a gap-fill task. Each target item was embedded in a supportive L2 sentence to prompt the meaning, and the participants were required to provide the base form of the target PVs. Each sentence contained two gaps, corresponding to each of the two component words (verb and particle) which formed the target PV. The first letter for each of the two words was provided:

> Now she feels the time has come to h___ o___ the business to someone else. (***To give power or control to someone else***).

At the end of each sentence, the meanings of the target PVs were given in brackets and printed in bold italics to make them more noticeable. This test was administered immediately after each learning session and again a week later.

A binary scoring system was used to score this test. The responses were scored as 1 when they were exactly the same as the target phrase or had minor spelling errors that did not make a response ambiguous (e.g., "*pit out*" or "*put aut*" instead of "*put out*"). If the response was ambiguous or wrong, it was scored as 0 (e.g., "*pay out*" instead of "*pay off*" or "*put off*" instead of "*pay off*"). Further, the responses were scored as 0 when no particle or verb was provided, or the response was not recognizable. After the first author had scored all posttests, 20% of the test responses randomly sampled from the data were scored by another researcher. A high interrater reliability of 91.88% based on the overall agreement was considered acceptable.

Meaning generation test

A meaning generation test was used to measure the participants' meaning recall of the PVs. The knowledge tested by the meaning generation test is similar to the type of knowledge required during reading (Elgort et al., 2020). In this pen-and-paper posttest, the target PVs were presented in weakly constraining sentences that did not provide strong support for guessing meaning from context (e.g., *They finally decided to* ***break up.***); there were no words in these sentences that revealed the meaning of the target item. The sentences were designed by the first author and checked by two other researchers. The participants were asked to write the PV meanings either in their L1 (Persian) or in L2 (English). This test was administered immediately after each learning session and again a week later.

A binary scoring system was used to score this test. If the response was the same or close to the dictionary translation or definition, it received a score of 1 (e.g., for *rip off* meaning "to cheat somebody by charging too much money for something," the response "to rob people's money" was considered a close definition), otherwise, it received a score of 0 (e.g., for *shake off* meaning "to escape," the response "to move backwards and forwards" was scored 0). After the first author had scored all posttests, 20% of the data were scored by an experienced researcher. This subsample was randomly selected to ensure representativeness. An interrater reliability of 92.26% was obtained, based on the overall percentage of agreement.

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

Test of prior knowledge

Although the PVs that were likely to be known by participants were not included prior to the data collection (based on the results of the norming studies), the participants also indicated their prior knowledge of the target PVs. After the MG posttest, the participants were presented with a list of the target PVs and asked to indicate if they knew the meaning of each item before the start of the study. They were instructed to choose 'Yes' if they knew the meaning and 'No' if they did not.

Experimental procedure

The information about the study was presented to the participants prior to the treatment. All participants read the information sheet, signed the consent form, and agreed to participate in this study. Then, they took the pretests. The learning phase started in three sessions. Figure 1 displays the overview of the experimental procedure in the learning phase for one class (the order of the PV set was counterbalanced across classes).

The participants were asked to read the texts and then to complete posttests. In each session, participants read two sets of texts. In Sessions 2 and 3, definitions were given before or after each text (so, if the text contained *break up*, the definition of *break up* was provided before or after that text). Each set comprised seven unique texts which were developed for the target PVs. Participants read one set of texts with unenhanced PVs, and the other set of texts with enhanced PVs. The learning phase was followed by the immediate posttests of the form and meaning of the target items. The same tests were

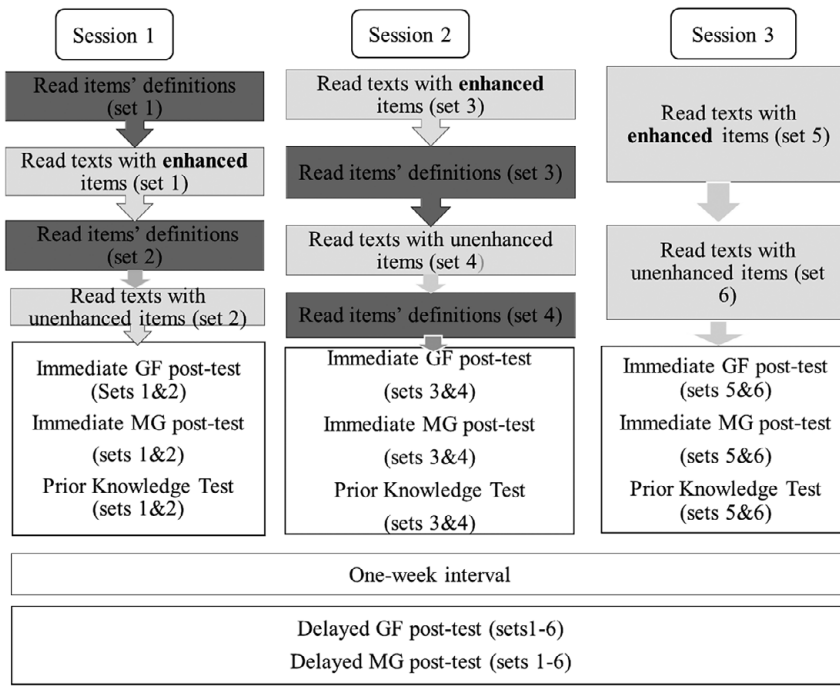

Figure 1. An overview of the experimental procedure for one class.

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

conducted as delayed posttests a week later. To ensure internal validity, the items were counterbalanced; the order of the sets within each group was also counterbalanced to counteract the sequencing effect (Appendix A).

Data analysis

Mixed effects models were fitted to the data, using the glmer function in the lme4 package in R (version 4.3.0, Bates, Mächler, Bolker, & Walker, 2015). Separate models were fitted to the two binary outcome variables: gap-fill posttest (GF) (correct = 1, incorrect = 0) and meaning generation posttest (MG) (correct = 1, incorrect = 0). Primary interest predictors were definition placement with three levels, typographic enhancement with two levels, and test time with two levels (immediate and delayed), and their interactions. The covariates were vocabulary levels test scores (numeric out of 100), and the following PV variables: word 1 frequency, phrase frequency, and prior knowledge of the PV. To interpret the interactions, Bonferroni-adjusted pairwise comparisons were conducted using the emmeans function in the emmeans package (version 1.4, Lenth, 2018). All continuous data was log-transformed and centered to avoid multicollinearity (Frost, 2014). After checking, Word 1 frequency was removed from the model, as it correlated highly with PV frequency ($r = .75$). Collinearity was checked using the VIF (Variance Inflation Factor) and the kappa coefficient. All categorical variables were contrast-coded (Brehm & Alday, 2022).

The initial models contained fixed effects for primary interest predictors, their interactions, and covariates. Both models included random intercepts for items and participants. A backward stepwise variable selection procedure was used to fit a minimally adequate statistical model to the data. The likelihood ratio test was used to compare models (Baayen, Davidson, & Bates, 2008). Vocabulary level test scores did not improve the model fit in either the MG or GF analysis ($\chi^2 = .001$, $p = .97$, $\chi^2 = 2.84$, $p = .09$, respectively). Fitting all random slopes for the primary interest predictors (including the interaction), caused convergence errors in both models. Therefore, the random slopes for the interaction terms were tested one at a time.

In the MG model, removing the interaction terms between definition placement and test time and definition placement and typographic enhancement negatively affected the model fit ($\chi^2 = 10.70$, $p < .001$ and $\chi^2 = 13.49$, $p = .001$, respectively). Therefore, these interactions were retained in the model. In the GF model, removing the interaction term between definition placement and test time negatively affected the model fit ($\chi^2 = 7.57$, $p = .02$). Although removing the interaction between definition placement and typographic enhancement ($\chi^2 = 5.42$, $p = .07$) did not significantly affect the model fit, the interaction was retained because it was of primary interest for the study.

Effects sizes were calculated as odds ratios, and standardized effect sizes (Cohen's d) were determined using the approach suggested by Chinn (2000). Effect sizes were interpreted following Brysbaert and Stevens (2018) suggesting that a typical effect size in similar psychology studies is between $d = .3$ and $d = .4$.

Results

Gap-fill test

Descriptive results of the proportion of correct responses on the immediate and delayed GF posttests are shown in Table 2. Presenting definitions (either before or after the text) led to higher GF scores than providing no definition, in both the immediate and delayed GF posttests.

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

Table 2. Means and 95% confidence intervals (in parentheses) of the immediate and delayed Gap-fill (GF) posttest

	No definition		Definition before		Definition after	
	TE = No	TE = Yes	TE = No	TE = Yes	TE = No	TE = Yes
Immediate	.18 (.15/.22)	.27 (.26/.30)	.74 (.70/.78)	.78 (.74/.82)	.78 (.75/.82)	.85 (.82/.88)
	.23(.21/.25)		.76(.73/.79)		.82(.79/.84)	
Delayed	.04 (.02/.05)	.07 (.05/.09)	.27 (.24/.29)	.34 (.31/.37)	.38 (.34/.41)	.54 (.50/.58)
	.05(.04/.06)		.30(.28/.33)		.46(.43/.49)	

On average, participants were able to achieve 60% form recall accuracy on the immediate posttest and 27% accuracy on the delayed posttest. Mixed-effects regression analysis showed that there were significant main effects of definition placement (χ^2 = 1110.29, p < .001), typographic enhancement (χ^2 = 63.64, p < .001), test time (χ^2 = 844.01, p < .001), prior knowledge (χ^2 = 216.73, p < .001), and PV frequency (χ^2 = 22.45, p < .001). There was a significant interaction between definition placement and test time (χ^2 = 7.57, p < .05), but the interaction between definition placement and typographic enhancement fell short of being statistically significant (χ^2 = 5.43, p = .07).

The results of the Bonferroni-adjusted pairwise post hoc comparisons for the interaction between definition placement and test time (Table 3) showed that the odds of getting a correct score in the GF posttest significantly increased from the condition where no definitions were available to seeing definitions before the texts, and even more so when definitions were presented after reading (Figure 2). The difference between before and after conditions was higher in the delayed than the immediate GF posttest (*OR* = 2.38, *OR* = 1.64, respectively).

The results also showed that there were main effects of prior knowledge and PV frequency, indicating that the PVs that had higher corpus frequency and were reported as known by the participants were recalled significantly more accurately in the GF posttest (p < .001). Although the effect of PV frequency was small (d = .20), the effect of prior knowledge can be considered medium (d = .47).

The meaning generation test

Descriptive results of the proportion of correct responses on the immediate and delayed MG posttests are presented in Table 4. Participants received the highest scores when the definitions were presented after the text and the target items were enhanced.

Table 3. Post hoc comparisons with Bonferroni-adjusted p-values for the interaction of definition placement and test time in the GF posttest

Comparisons		Δ of probability of response accuracy	*SE*	*Odds ratio*	*z*	*p*
After - before	Test = Immediate	.04	.18	1.64	4.59	<.001
	Test = Delayed	.21	.22	2.38	9.21	<.001
After - none	Test = Immediate	.63	3.60	30.77	29.29	<.001
	Test = Delayed	.59	5.26	32.80	21.78	<.001
Before - none	Test = Immediate	.59	2.06	18.79	26.73	<.001
	Test = Delayed	.37	2.21	13.80	16.42	<.001

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

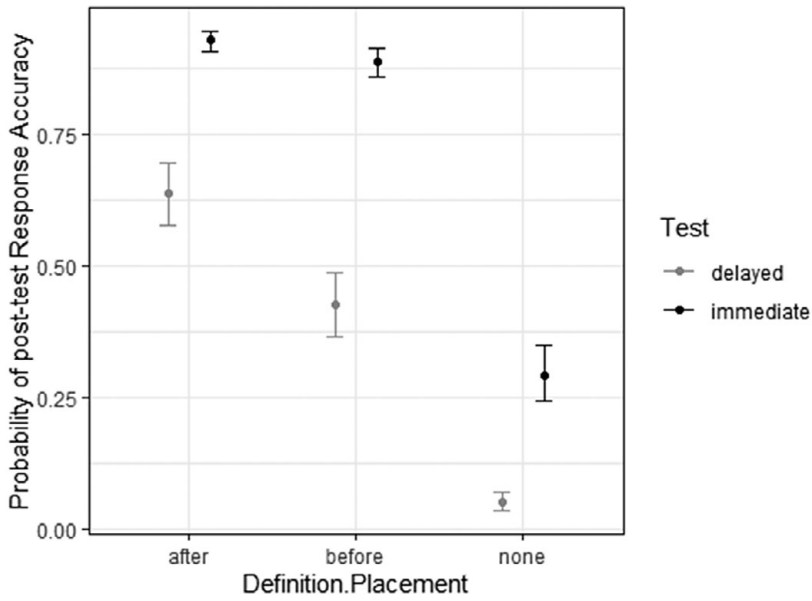

Figure 2. Estimated interaction between definition placement and test time in the GF posttest.

Table 4. Means and 95% confidence intervals (in parentheses) of the immediate and delayed meaning generation (MG) posttests

	No definition		Definition before		Definition after	
	TE = No	TE = Yes	TE = No	TE = Yes	TE = No	TE = Yes
Immediate	.16 (.13/.17)	.28 (.24/.32)	.76 (.73/.79)	.77 (.73/ .81)	.79 (.76/ .82)	.88 (.86/.90)
	.22 (.20/.24)		.76 (.74/ .79)		.83 (.81/.85)	
Delayed	.04 (.02/.05)	.07 (.05/.08)	.26 (.23/.30)	.34 (.30/.38)	.39 (.35/.42)	.57 (.54/.61)
	.05 (.04/.06)		.30 (.26/.33)		.48 (.45/.51)	

Note: TE - typographic enhancement

On average, participants were able to achieve 61% meaning recall accuracy on the immediate posttest and 28% accuracy on the delayed posttest. Mixed-effects regression analysis showed significant main effects of definition placement ($\chi^2 = 1180.38$, $p < .001$), typographic enhancement ($\chi^2 = 81.28$, $p < .001$), test time ($\chi^2 = 846.52$, $p < .001$), prior knowledge ($\chi^2 = 86.32$, $p < .001$), and PV frequency ($\chi^2 = 22.00$, $p < .001$). There were also significant interactions between definition placement and typographic enhancement ($\chi^2 = 13.12$, $p < .01$) and definition placement and test time ($\chi^2 = 10.72$, $p < .01$).

The results of the Bonferroni-adjusted pairwise post hoc comparisons for the interaction between definition placement and typographic enhancement (Table 5) showed that the odds of getting a correct score in the MG posttest significantly increased from the condition where no definitions were available to having access to definitions before the texts, and even more so when definitions were presented after reading (Figure 3). The difference between before and after conditions was higher when typographic enhancement was used (*OR* = 2.80, *OR* = 1.90, respectively).

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

Table 5. Post hoc comparisons with Bonferroni-adjusted *p*-values for the interaction between definition placement (before/after) and typographic enhancement (yes/no) in the MG posttest

Comparisons	TE	Δ of probability of response accuracy	*SE*	*Odds ratios*	*z*	*p*
After - Before	Yes	.18	.30	2.80	9.68	<.001
	No	.14	.19	1.90	6.33	<.001
After - None	Yes	.71	4.41	34.20	27.35	<.001
	No	.68	5.85	40.8	25.85	<.001
Before - None	Yes	.53	1.50	12.20	20.29	<.001
	No	.54	3.03	21.50	21.78	<.001

Note: TE - typographic enhancement

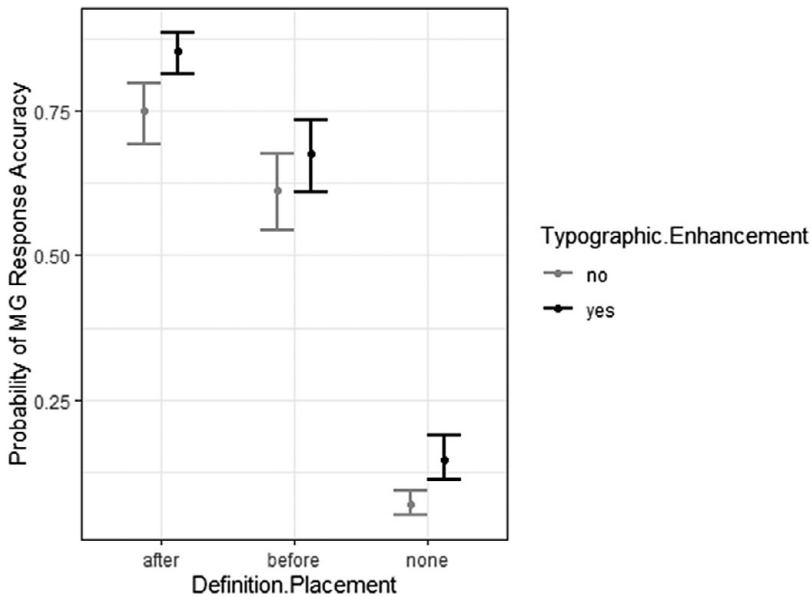

Figure 3. Estimated interaction between typographic enhancement and definition placement in the MG posttest.

The results of the Bonferroni-adjusted pairwise post hoc comparisons for the interaction between definition placement and test time (Table 6) showed that the odds of getting a correct score in the MG posttest significantly increased from the condition where no definitions were available to seeing definitions before the texts, and even more so when definitions were presented after reading (Figure 4). The difference between before and after conditions was higher in the delayed than in the immediate MG posttest (*OR* = 2.81, *OR* = 1.89, respectively).

There were also main effects of prior knowledge and PV frequency ($p < .001$; $d = .29$, $d = .23$, respectively) in the MG analysis. This indicates that the accuracy of meaning recall was significantly higher for higher frequency PVs and the PVs considered known by participants.

General discussion

This study investigated the effects of definition placement and typographic enhancement on the contextual learning and retention of PVs. In total, 85 high school students

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

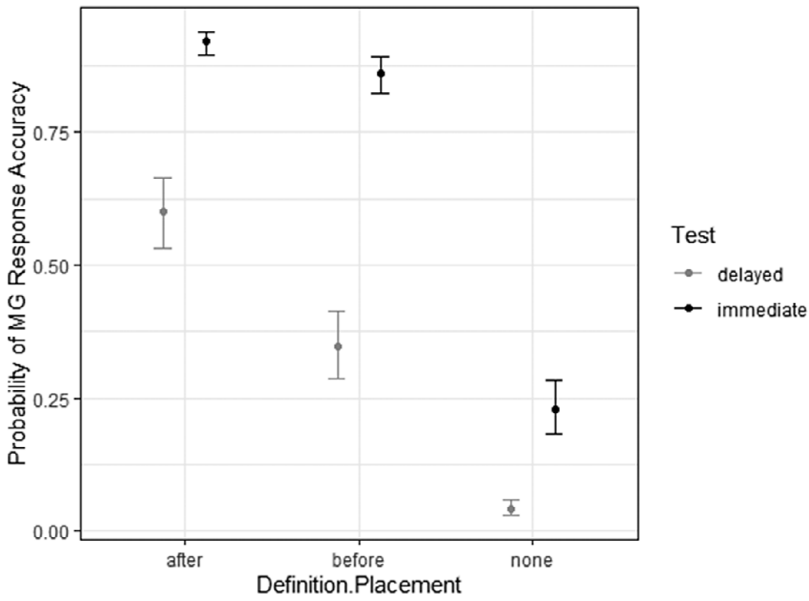

Figure 4. Estimated interaction between definition placement and test time in the MG posttest.

Table 6. Post hoc comparisons with Bonferroni-adjusted p-values for the interaction of definition placement and test time in the MG posttest

Comparisons	Test	Δ of probability of response accuracy	*SE*	*Odds ratios*	*z*	*p*
After - before	Immediate	.06	.21	1.89	5.72	<.001
	Delayed	.25	.27	2.81	10.85	<.001
After - none	Immediate	.69	4.74	39.11	30.26	<.001
	Delayed	.56	5.69	35.62	22.36	<.001
Before - none	Immediate	.63	2.32	20.71	27.09	<.001
	Delayed	.31	2.01	12.66	16.01	<.001

read 42 texts in which the target items were repeated three times, and their immediate and delayed knowledge was measured using a gap-fill and a meaning generation posttest. It was found that on average (across MG and GF posttests), participants recalled 60.5% and 27.5% of the target items in the immediate and delayed measurements, respectively. The results showed that when no definition was given, participants recalled, on average (across two TE conditions), about 23% of the target PVs in the immediate and only 5% in the delayed measurements (Tables 2 and 4), which aligns with previous findings showing that contextual learning of L2 MWEs without support is relatively weak (e.g., Szudarski, 2012; Szudarski & Carter, 2016: Webb et al., 2023). However, even this amount of learning without definitions is surprising, given that the texts in the present study did not strongly constrain the meaning of the target PVs. The research questions posed in the present study are addressed below.

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

RQ1: Is contextual learning and retention of PVs from reading affected by the provision of definitions?

The answer to RQ1 is yes. The instance-based memory model of vocabulary learning predicts that providing definitions should increase contextual vocabulary learning and retention (Reichle & Perfetti, 2003). The results of the present study support this prediction. Due to the figurative nature of the target PVs, students could not readily guess the meaning of the whole phrase from its constituent parts. In the absence of strong contextual clues, their meaning inferences during reading were likely incomplete. Providing definitions of the PVs facilitated their contextual learning because the core semantic features of the PVs' figurative meanings in the definitions likely resonated with the correct contextual meaning inferences, facilitating the establishment of PVs' semantic representations (Bolger et al., 2008). Without definitions, little initial contextual PV learning and almost no retention were observed (see Figures 2 and 4).

Because presenting PVs with their definitions improved both form and meaning recall of the target PVs, access to definitions likely facilitated the form-meaning mapping for the target PVs by explicitly communicating the correct figurative meanings of the PVs. The results of the current study further corroborate previous empirical findings showing that providing definitions improves contextual vocabulary learning (AbuSeileek, 2011; Bolger et al., 2008; Elgort et al. 2020; Hulstijn et al., 1996).

RQ2: Is contextual learning and retention of PVs from reading affected by definition placement?

The answer to RQ2 is yes. This finding is in line with the lexical inferencing theory (Bertsch et al., 2007) and the semantic elaboration model (Craik & Tulving, 1975). Inferring the meanings of PVs during reading before accessing their definitions is predicted to lead to greater learning and retention of PVs than previewing definitions before reading. Indeed, we found that when contextual inferences are followed by correct definitions, the learning, and retention of the target PVs were better compared with the learning condition in which contextual inferences followed definitions (i.e., in the definition-before-text condition). These results are in line with previous learning and memory research findings indicating that inferencing enhances learning as it requires a greater degree of cognitive effort compared with conditions in which there was no opportunity to infer the meanings (e.g., Bertsch et al., 2007; Bjork & Bjork, 2014). Furthermore, inference-making in context is a form of semantic elaboration, which has been shown to lead to durable memory traces (Craik & Tulving, 1975).

The results are also in line with L2 vocabulary research that found presenting definitions after reading was an effective method for learning L2 words (e.g., Elgort et al., 2020; Huang & Lin, 2014). This finding contrasts with the results reported by Strong and Boers (2019) that, when learners engage in blind guessing (e.g., in gap-fill textbook exercises with MWEs), incorrect responses create erroneous memory traces that prevent the learning of correct MWEs. However, incorrect meaning inferences in reading are different from gap-fill exercises; for one, in contextual learning from reading, MWEs are presented as intact phrases, creating accurate representations of the whole MWEs. The findings may also suggest that even unsuccessful inferences followed by feedback (definitions) are better than presenting definitions before reading.

Indeed, we found that PV knowledge retention (measured by the delayed posttests) was better when participants first encountered target PVs in reading and then reviewed their definitions (see Figures 3 and 4). In the preview condition, we observed the largest

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

knowledge attrition between the immediate and delayed posttest. One explanation is that deeper word encoding occurred in the postview condition than in the preview condition (Rodriguez-Fornells, Kofidis, & Münte, 2004). Presenting definitions prior to reading familiarized students with the PVs and their meanings. Such familiarity may reduce attention allocation to the PV during reading (e.g., Elgort et al., 2020, 2023; Koriat & Bjork, 2005; Yang, Potts, & Shanks, 2017). In an eye-tracking study, for example, Elgort et al. (2023) found that the pseudowords that were previewed with definitions before reading were more likely to be skipped and were fixated on for a shorter time during reading. They attributed this behavior to the readers' perceived familiarity with the target items. These findings may explain why students performed better in the condition where definitions were provided after the text. Previewing definitions prior to reading may also discourage learners from making contextual inferences that are associated with deep encoding (Rodrigues-Fornells et al., 2004).

The results further showed that the interaction between definition placement and test time was significant in both posttests. In the immediate MG posttest, when definitions were previewed prior to reading, participants scored 76%, but when definitions were given after reading, they achieved 83% (an increase of 7%). In the delayed MG posttest, when definitions were presented prior to reading, participants' response accuracy was 30% but when definitions were presented after reading participants' response accuracy was 48% (an increase of 18%). The same pattern occurred in the immediate and delayed GF posttests; the difference between preview and postview conditions was greater in the delayed than in the immediate posttests. In summary, the effects of definition placement were greater in the delayed than immediate posttests.

RQ3: Is contextual learning and retention of PVs from reading affected by using typographic enhancement?

The answer to RQ3 is also yes. According to the noticing hypothesis, what learners either deliberately attend to or unintentionally notice in the input is more likely to become intake (Schmidt, 2001). Participants' superior performance in the enhanced condition suggests that employing typographic enhancement (bolding) did increase learners' attention to the target PVs in the present study, possibly, by making them more visually salient. Previous studies also showed that using enhancement techniques increased students' attention to the vocabulary items (e.g., Boers et al., 2017; Durrant & Schmitt, 2010; Puimège et al., 2023; Szudarski & Cartet, 2016; Webb et al., 2013).

The present findings further showed that typographic enhancement facilitated explicit knowledge of both the meaning and form of the critical PVs, which corroborates the findings of Sonbul and Schmitt (2013) and Toomer and Elgort (2019). These authors found that typographic enhancement of the L2 target collocations in reading texts positively influenced their learning, measured by form recall and form recognition tests.

These results contrast with the studies that found that typographic enhancement of figurative MWEs in reading and audiovisual materials did not improve their learning (Campillo, 2015; El-Dakhs et al., 2021; Majuddin, Siyanova-Chanturia, & Boers, 2021). Such differences may be due to the difference between the study materials, the type of item examined, and participants' English language proficiency. For example, Majuddin et al. (2021) used audiovisual input (an episode of a comedy series) which might have differentially affected students' level of engagement with the text. They also conducted their study with higher proficiency learners than those used in the present study.

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

Interestingly, they found that the impact of typographic enhancement was weaker for participants with higher vocabulary knowledge. This might suggest that advanced learners rely less on visual cues in subtitles during video watching, instead utilizing their existing vocabulary knowledge to comprehend figurative expressions.

RQ4: Does the effect of typographic enhancement modulate the effect of definition placement?

We conjectured that typographic enhancement would direct learners' attention towards PVs during reading, facilitating the noticing of the PVs. Typographic enhancement is therefore likely to be more beneficial in the condition where participants need to make contextual inferences about the PVs before being presented with the whole target phrase and its figurative meaning, compared with the condition where PVs and their definitions are presented prior to reading. This was the case in the present study. When typographic enhancement was not used, and definitions were given after reading, participants recalled 79% and 39% of target PVs in the immediate and delayed MG posttests, which increased to 88% and 57% when typographic enhancement was used.

Therefore, the answer to RQ4 is also yes for the development of the knowledge of meaning. Typographic enhancement appears to have increased students' attention to the items, whereas presenting definitions after the text prompted participants to make an inference regarding the potential meanings of the items. These inferences were then verified (if they were correct) or refined (if they were incorrect) once the correct meanings through definitions were subsequently provided after this stage. This explanation is supported by the findings of Elgort et al. (2023) who found that readers spent more time on items whose definitions were given after reading.

In contrast to the findings of the current study, however, Toomer and Elgort (2019) found that providing definitions did not create a collocation learning advantage over and above that of typographic enhancement. There are, however, some important differences between the studies. Definitions in Toomer and Elgort (2019) were given as in-text glosses while, in the present study, definitions were given either prior to reading or after learners finished reading. When definitions are provided in the margins or the text, they might interfere with the flow of reading and may negatively affect online processing. This disruption may have negatively affected the encoding of collocations as whole phrases. Furthermore, in Toomer and Elgort (2019), definitions were provided from the first time the target MWE occurred in the texts and were presented every time it occurred. This may have reduced the need for learners to generate contextual inferences about the meanings of the target MWEs in their study, negatively affecting engagement with the collocations in context. Also, they used fairly transparent lexical collocations (e.g., *partial response*) while the present study used figurative phrasal verbs. PVs consist of a verb and a particle, similar to grammatical collocations which consist of a content word and a preposition. Toomer et al. (2024) who investigated the effects of typographic enhancement on contextual learning of lexical (verb + noun) and grammatical (preposition + noun) collocations found that typographic enhancement was more effective for the learning of grammatical collocations. They argued that using typographic enhancement may have made grammatical collocations more perceptually salient as a whole expression during reading for the learners.

The present study, while contributing valuable insights into the effects of definition placement and typographic enhancement, is not without limitations. Completing the GF posttest prior to the MG posttest might have had an impact on the performance on

https://doi.org/10.1017/S0272263124000718 Published online by Cambridge University Press

the MG posttest. Participants saw the meanings of the items in brackets in the GF posttest, although the PVs themselves were not presented. This could have helped them remember the possible meanings of all PVs in the MG posttest (however, without an association between the specific meaning and their corresponding PV forms). As a reviewer recommended, one way to potentially minimize this effect could have been to use distractor items in the posttests.

Another limitation was the administration of the prior knowledge test before the delayed posttests. A reviewer pointed out that this could have made the PVs more salient to the participants and helped them on the delayed Gap-fill posttest, and we agree. However, the intact form of the PVs was also presented to the participants during the learning phase, at the time when they were introduced to their definitions. These two limitations would have affected the experimental manipulations (i.e., presentation of the definitions, presence or absence of typographic enhancement) similarly. Finally, some variability in figurativeness might have affected the learning of the target PVs. However, because the items selected for the study had been rated as most figurative, with narrow figurativeness ranges, it is unlikely that this variability would have significantly affected the present findings.

Conclusions and implications

The present study contributes to the line of research that looks for optimal conditions for learning L2 MWEs, vocabulary items that are known to be challenging for language learners. First, we aimed to test whether providing definitions of novel vocabulary to supplement contextual learning would be as effective for MWE learning as it is for learning single words. Second, following Elgort et al. (2020), we set out to test whether definitions should be accessed before or after encountering novel L2 vocabulary in reading (i.e., before or after making contextual inferences). Third, we tested whether typographic enhancement provides an additional learning boost when readers encounter these MWEs in the text, especially when these expressions have not been introduced prior to reading.

To do so, the impacts of definition placement, typographic enhancement, and their interaction on contextual learning and retention of 42 English figurative PVs were tested with 85 learners of English. The results showed a clear advantage of learning with definitions and presenting definitions after reading. It was further found that employing typographic enhancement increased the learning advantage of accessing definitions and reading for the knowledge of PV meanings. Based on these findings, we conclude that the use of definitions should be encouraged in contextual learning of figurative MWEs. Learners' contextual learning and retention of MWEs will further benefit from inferring meanings of novel MWEs during reading, prior to consulting definitions. Our advice to teachers, publishers, and material developers is not to take inference opportunities away from L2 learners for fear of possible erroneous contextual inferences. An analysis of English learning textbooks showed that learners do not have sufficient opportunities to learn MWEs effectively from input (Strong & Boers, 2019). The results of the present study clearly show that supplementing reading texts with definitions presented after reading can provide such input. The use of typographic enhancement on target MWEs in the reading text is likely to further improve their learning.

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at https://url.avanan.click/v2/r02/.

References

AbuSeileek, A. F. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education*, *57*(1), 1281–1291. https://doi.org/10.1016/j.compedu.2011.01.011

Anthony, L. (2022). *AntWordProfiler* (Version 2.0.0) [Computer Software]. Tokyo, Japan: Waseda University. http://www.antlab.sci.waseda.ac.jp/

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher and J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 59–68). Worth.

Boers, F. (2020). Factors affecting the learning of multiword items. In Stuart Webb (ed.), *The Routledge handbook of vocabulary studies* (pp. 143–157). Routledge.

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*, 448–469. https://doi.org/10.1111/ijal.12141

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & M. Demecheleer (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*(3), 245–61. https://doi.org/10.1191/1362168806lr195oa

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Contextual variation and definitions in learning the meaning of words. *Discourse Processes*, *45*, 122–159. https://doi.org/10.1080/01638530701792826

Brehm, L., & Alday, P.M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*(4). https://doi.org/10.1016/j.jml.2022.104334

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Campillo, P., S. (2015). Effect of textual enhancement on idioms: An exploratory study with Spanish students. *Revista Española de Lingüística Aplicada*, *28*(1), 258–272. https://doi.org/10.1075/resla.28.1.12sal

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*(22). https://doiorg.helicon.vuw.ac.nz/10.1002/10970258(20001130)19:22%3C3127::AID-SIM784%3E3.0.CO;2-M

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3). 403–426.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294. http://doi.org/10.1037/0096-3445.104.3.268

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

Eguchi, A., & Kyle, K. (2023). L2 collocation profiles and their relationship with vocabulary proficiency: A learner corpus approach. *Journal of Second Language Writing*, *60*. https://doi.org/10.1016/j.jslw.2023.100975

El-Dakhs, D., Sonbul, S., & Alwazzan, R. (2021). Learning phrasal verbs in the EFL classroom: the effect of prior vocabulary knowledge and opacity. *International Review of Applied Linguistics in Language Teaching*, *60*(4), 1253–1291. https://doi.org/10.1515/iral-2020.0116

Elgort, I., Beliaeva, N., & Boers, F. (2020). Contextual word learning in the first and second language: Definition placement and inference error effects on declarative and nondeclarative knowledge. *Studies in Second Language Acquisition*, *42*(1), 7c32. https://doi.org/10.1017/S0272263119000561

Elgort, I., van de Wetering, R., Arrow, T., & Beyersmann, E. (2023). Previewing novel words before reading affects their processing during reading: An eye-movement study with first and second language readers. *Language Learning*, *74*(1), 78. https://doi.org/10.1111/lang.12579110; https://doi.org/10.1111/lang.12579

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*(2), 365–414. https://doi.org/10.1111/lang.12052

Frost. J. (2014). *Multicollinearity in Regression Analysis: Problems, Detection, and Solutions*. Statistics By Jim. https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*, 339–360. https://doi.org/10.1016/j.system.2016.04.004

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

Huang, L. L., & Lin, C. C. (2014). Three approaches to glossing and their effects on vocabulary learning. *System*, *44*. 127–136. https://doi.org/10.1016/j.system.2014.03.006

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*, 327–339. https://doi.org/10.2307/329439

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). *Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel*. Research Branch Report, 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology:* Learning, Memory, and Cognition, *31*, 187–194. http://doi.org/10.1037/0278-7393.31.2.187

Lenth, R. (2018). *Estimated marginal means, aka least-squares means*. https://github.com/rvlenth/emmeans

McCarthy, M., & O'Dell, F. (2004). *English phrasal verbs in use*. Cambridge University Press.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions from audiovisual input: the role of repetition and typographic enhancement *Studies in Second Language Acquisition*, *43*(5), 985–1008. https://doi.org/10.1017/S0272263121000036

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

Northbrook, J., Allen, D., & Conklin, K. (2022). 'Did you see that?'—The role of repetition and enhancement on lexical bundle processing in English learning materials. *Applied Linguistics*, *43*(3), 453–472, https://doi.org/10.1093/applin/amab063

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning from listening to L2 songs. *Studies in Second Language Acquisition*, *41*(4), 745–768. https://doi.org/10.1017/S0272263119000020

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*(1), 162–203. https://doi.org/10.1111/lang.12430

Peters, E. (2012). Learning German formulaic sequences: the effect of two attention-drawing techniques. *The Language Learning Journal*, *40*(1), 65–79. https://doi.org/10.1080/09571736.2012.658224

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438. https://doi.org/10.1111/lang.12364

Puimège, E., Montero Perez, M., & Peters, E (2023). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, *39*(2). 471–492. https://doi.org/10.1177/02676583211049741

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*, 219–237. https://doi.org/10.1207/S1532799XSSR07032

Rodriguez-Fornells, A., Kofidis, C., & Münte, T.F. (2004). An electrophysiological study of errorless learning, *Cognitive Brain Research*, *19*(2), 160–173, https://doi.org/10.1016/j.cogbrainres.2003.11.009

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381–402. https://doiorg.helicon.vuw.ac.nz/10.1177/1362168815618428

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp.3–32). Cambridge University Press.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi-org.helicon.vuw.ac.nz/10.1177/026553220101800103

Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. Routledge.

Sonbul, S., & Schmitt, R. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Strong, B., & Boers, F. (2019). Weighing up exercises on phrasal verbs: Retrieval versus trial-and-error practice. *The Modern Language Journal*, *103*(3). 562–579. https://doi.org/10.1111/modl.12579

Szudarski, P. (2012). Effects of meaning- and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching and Research*, *1*(2) 3–37. https://api.semanticscholar.org/CorpusID:220740526

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2), 245–265. https://doi.org/10.1111/ijal.12092

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, *69*, 405–439. https://doi.org/10.1111/lang.12335

Toomer, M., Elgort, I., & Coxhead, A. (2024). Contextual learning of L2 lexical and grammatical collocations with and without typographic enhancement. *System*, *121*. https://doi.org/10.1016/j.system.2024.103235

Vilkaitė, L, & Schmitt, N. (2017). Reading Collocations in an L2: Do Collocation Processing Benefits Extend to Non-Adjacent Collocations? *Applied Linguistics 40*(2). 329–354. https://doi.org/10.1093/applin/amx030

Vu, D. V., & Peters, E. (2022). Incidental learning of collocations from meaningful input: A longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, *44*(3), 685–707. doi:10.1017/S0272263121000462

Wang, X., Boers, F., & Warren, P. (2022). Prompting language learners to guess the meaning of idioms: Do wrong guesses linger? *Language Awareness*, *33*(1) 94–116. https://doi.org/10.1080/09658416.2022.2153859

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, *168*(1), 34–70. https://doi.org/10.1075/itl.168.1.02web

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, *56*(2), 161–180. https://doi.org/10.1017/S0261444822000507

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, *35*, 182–191. https://doi.org/10.1016/j.system.2006.12.009.

Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1073–1092. http://doi.org/10.1037/xlm0000363

Yu, X., & Boers, F. (2023). Inferring the meaning of idioms: Does accuracy matter for retention in memory? *RELC Journal*, *55*(3), 721–734. https://doi.org/10.1177/00336882231181771

Zhang, C., & Ma, R. (2021). The effect of textual glosses on L2 vocabulary acquisition: A meta-analysis. *Language Teaching Research*, *28*(3), 967–986. https://doi.org/10.1177/13621688211011511

Appendix A

Counterbalancing of experimental conditions

Group 1	Sets and items	Definition placement	Typographic enhancement	Condition
	Set 1 (1, 2, 3, 4, 5, 6, 7)	Before	Yes	1
	Set 2 (8, 9, 10, 11, 12, 13, 14)	Before	No	2
	Set 3 (15, 16, 17, 18, 19, 20, 21)	After	Yes	3
	Set 4 (22, 23, 24, 25, 26, 27, 28)	After	No	4
	Set 5 (29, 30, 31, 32, 33, 34, 35)	None	Yes	5
	Set 6 (36, 37, 38, 39, 40, 41, 42)	None	No	6
Group 2				
	Set 2 (8, 9, 10, 11, 12, 13, 14)	Before	Yes	1
	Set 3 (15, 16, 17, 18, 19, 20, 21)	Before	No	2
	Set 4 (22, 23, 24, 25, 26, 27, 28)	After	Yes	3
	Set 5 (29, 30, 31, 32, 33, 34, 35)	After	No	4
	Set 6 (36, 37, 38, 39, 40, 41, 42)	None	Yes	5
	Set 1 (1, 2, 3, 4, 5, 6, 7)	None	No	6
Group 3				
	Set 3 (15, 16, 17, 18, 19, 20, 21)	Before	Yes	1
	Set 4 (22, 23, 24, 25, 26, 27, 28)	Before	No	2
	Set 5 (29, 30, 31, 32, 33, 34, 35)	After	Yes	3
	Set 6 (36, 37, 38, 39, 40, 41, 42)	After	No	4
	Set 1 (1, 2, 3, 4, 5, 6, 7)	None	Yes	5
	Set 2 (8, 9, 10, 11, 12, 13, 14)	None	No	6
Group 4				
	Set 4 (22, 23, 24, 25, 26, 27, 28)	Before	Yes	1
	Set 5 (29, 30, 31, 32, 33, 34, 35)	Before	No	2
	Set 6 (36, 37, 38, 39, 40, 41, 42)	After	Yes	3
	Set 1 (1, 2, 3, 4, 5, 6, 7)	After	No	4
	Set 2 (8, 9, 10, 11, 12, 13, 14)	None	Yes	5
	Set 3 (15, 16, 17, 18, 19, 20, 21)	None	No	6
Group 5				
	Set 5 (29, 30, 31, 32, 33, 34, 35)	Before	Yes	1
	Set 6 (36, 37, 38, 39, 40, 41, 42)	Before	No	2
	Set 1 (1, 2, 3, 4, 5, 6, 7)	After	Yes	3
	Set 2 (8, 9, 10, 11, 12, 13, 14)	After	No	4
	Set 3 (15, 16, 17, 18, 19, 20, 21)	None	Yes	5
	Set 4 (22, 23, 24, 25, 26, 27, 28)	None	No	6

Cite this article: Tadayonifar, M., Elgort, I., & Siyanova-Chanturia, A. (2025). Contextual learning and retention of phrasal verbs: The effects of definition placement and typographic enhancement. *Studies in Second Language Acquisition,* 47: 157–180. https://doi.org/10.1017/S0272263124000718