The Modern Language Journal

LITERATURE AND SECOND LANGUAGE VOCABULARY LEARNING: THE ROLE OF TEXT TYPE AND TEACHING APPROACH

Journal:	*The Modern Language Journal*
Manuscript ID	MLJ-23-0148-04.R1
Wiley - Manuscript type:	Original Article
Keywords:	literature, poetry, vocabulary, reading, French

LITERATURE AND SECOND LANGUAGE VOCABULARY LEARNING: THE ROLE OF TEXT TYPE AND TEACHING APPROACH

ABSTRACT

This study considers the relative benefits for vocabulary learning of exposure to two types of texts, literary or non-literary, used with two teaching approaches. These approaches were termed *functional* and *creative* respectively. In the former, learners' attention was drawn to factual information and linguistic features in order to develop their linguistic knowledge. In the latter, the aim was to stimulate learners' personal and emotional response, by drawing their attention to the text's emotional content and how language was used to express meaning. We analysed data from 160 learners of French in eight schools in England. Learners in four schools studied French poems and those in another four studied French factual texts. Teachers in each text condition employed *functional* and *creative* methods of exploitation within a counterbalanced design. We assessed two types of vocabulary knowledge at pre- and post-test: meaning-recall of vocabulary contained in the texts, and learners' general vocabulary size. Our results indicated learning gains across both text-types. There were however important interactions between text-type and teaching approach and between text-type and the order in which the teaching approaches were used. Finally, we consider the implications of these findings for understanding of vocabulary learning through literature and for classroom practice.

Keywords: literature; poetry; vocabulary; reading; French

INTRODUCTION

This study investigated the impact of exposure to French poems on the vocabulary development of high school learners of French, compared with the impact of non-literary texts. It was motivated by the fact that, in many educational contexts, the use of literary texts

is viewed as a key part of the language curriculum. This is equally true of England, the context of the present study. Nevertheless, the impact of using such materials on school-aged learners' language development has received little research attention, as most studies have focused on adult language development (e.g. Kim, 2004; O'Donnell, 2009; Yang, 2001), or on affective responses to literature (Paran, 2008) rather than on its impact on language acquisition. While affective responses are of course important, in school settings linguistic outcomes are likely the prime focus of teachers and learners alike. In England, as elsewhere, improving learners' vocabulary knowledge is a key concern (DfE, 2022). In this study, therefore, our first objective was to compare vocabulary learning gains from reading literary texts with gains from factual texts. To our knowledge, this is the first study to make a direct comparison between those two text-types with younger, school-based language learners.

Our second objective was to explore whether the pedagogical approach used with these two text-types affected their impact on vocabulary learning. This is important because while much adult reading of foreign language literature may take place outside of the classroom in a fairly-self-directed manner, in school settings it is most likely to occur under the close direction of the teacher. While there is agreement that **how** literary texts are used is crucial (Paran, 2008), there is an ongoing debate about the most appropriate form of deployment. We are not aware of previous studies that have explored empirically the impact of teaching approach on outcomes from studying literary texts, nor any that have examined both text-type and teaching approach in one investigation. Understanding the impact of both factors on language outcomes is essential not only to guide pedagogical and curricular decision-making, but also to enhance understanding of vocabulary learning through literature on a theoretical level. The study therefore makes a novel and significant contribution not only to the under-researched area of literature and school-based language learning but also to the field of instructed vocabulary learning more broadly.

LITERATURE REVIEW

Pedagogical models for using literature in language teaching

How literature should be used in language classrooms (rather than as extensive reading outside the classroom, as for example, in Pellicer-Sánchez & Schmitt, 2010) has been the subject of discussion for a number of years. Different models of, or approaches to, literature and language teaching are outlined in Bobkina and Dominguez (2014) and elsewhere (e.g. Bloemert et al., 2016). While these models are varied, two broadly contrasting teaching approaches can be extracted from them: those in which the focus is primarily on developing learners' linguistic knowledge and skills, and those in which more person-focused development is the aim, such as cultural knowledge or the readers' personal involvement with the text. For example, in the Personal Growth Model or approach (henceforth, approach) outlined in Carter and Long (1991) alongside a Language Model and a Culture Model, learners undertake activities through which they relate the text to their personal experiences and express their feelings and emotions about what they have read. Learners themselves also seem to be aware of a distinction between different approaches to studying literature; for example, Tsang et al. (2020) found that high school learners' perceptions of the benefits of studying literature crystallised around language development and personal development respectively. Of course, the two are interrelated; to engage on a personal level learners need to draw on their linguistic knowledge, especially vocabulary knowledge, opening up the possibility of incidental learning occurring. How much vocabulary is learnt from literary texts compared with factual texts, and whether such learning varies according to teaching approach used, are issues which to our knowledge have not been explored in research previously.

A rare study that partly explores those questions, Maxim (2002), examined how much vocabulary knowledge can be gained through reading literature and how that compared with vocabulary learning from non-literary material for university beginner learners of German in

the US over 10 weeks. Learners in a literature group read a novel in class which was at a linguistic level above their proficiency level. Classroom activities used with the novel included story re-tell, text analysis and interpretation. Although not explicitly stated by Maxim, these activities seemed to have had some focus on personal and cultural development, alongside offering opportunities for language enhancement. A comparison group read passages from the textbook in order to practise vocabulary and grammar. Both groups completed vocabulary tests before and after the intervention. These required learners to give English definitions for selected vocabulary items in reading passages (called a 'Definition' task) and then make some sort of evaluation about each item, for example, whether it had a positive or negative association or what its function in a sentence was (termed the 'Functional' task). There were no significant differences in how much progress each group made; substantial vocabulary learning occurred for both, although the literature group made the most progress on the Functional task and the textbook group on the Definition task. This raises two questions: whether different text-types might lead to the development of different kinds of vocabulary knowledge, and whether it was the literary text, or the teaching approach used, that was most responsible for the vocabulary gains seen in Maxim (2002). To better understand those issues, theories of second language (L2) vocabulary learning in instructed settings need to be considered, which we turn to next.

VOCABULARY LEARNING IN INSTRUCTED CONTEXTS: TEXT-TYPE AND TEACHING APPROACHES

The learning of vocabulary in instructed second language contexts is enhanced, according to Schmitt (2008), by "virtually anything that leads to more exposure, attention, manipulation, or time spent on lexical items"; in other words, by anything that brings greater "engagement" (p. 339). Definitions of 'engagement' are varied but in relation to L2 learning

it has been conceived of as "a state of heightened attention and involvement" which includes interacting cognitive, behavioural and emotional aspects (Philp & Duchesne, 2016: 51). In this section we consider engagement first from the perspective of 'attention' in relation to different text-types. We then consider how different teaching approaches or learning activities might lead to different levels of 'involvement' on the part of learners.

Text-type: attention and noticing

Formulated originally to account for the acquisition of grammar, Schmidt's (1990) Noticing Hypothesis has also been drawn upon to argue that for vocabulary learning to occur, learners have to 'notice' or 'attend to' words within the input they receive. That may be especially the case for incidental learning, where words are learnt as a byproduct of an activity where the focus is on meaning comprehension or communication (Feng & Webb, 2020; Laufer & Hulstijn, 2001; Pouresmaeil & Vali, 2023). Literary texts may facilitate such noticing, because one of their central features, especially in highly stylised genres such as poetry, is that language may be used in an unusual, unfamiliar and potentially ambiguous way. The reader is thus prompted to notice the gap between that novel deployment and the meaning they have previously attributed to such language (Johnson-Laird & Oatley, 2022). In such instances learners have to work harder to search for the possible meaning of the word in a novel and slightly non-transparent context, causing their reading and comprehension to slow down and potentially better learning to occur (Lui et al., 2021). For second language learners, however, especially of a lower proficiency level, it is likely to be important that comprehension is not slowed down to such an extent that limited learning occurs, and that a level of 'desirable difficulty' is presented to them (Bjork & Bjork, 2011).

Poetry thus offers readers input that "is foregrounded, unusual, and draws attention to itself" (Hanauer, 2001: 298). Furthermore, attending to form-meaning connections is crucial for language acquisition (Schmidt, 2001). In literature, form and meaning are closely

integrated – the writer uses language forms to convey a specific meaning, giving learners potentially a more nuanced understanding of language and its use (Paesani et al., 2015). In that respect, poetry stands out because poems can draw learners' attention to formal aspects of language, including, for example, phonology, "within a meaning construction framework" (Hanauer, 2001: 298), with increased potential for better and richer word learning of vocabulary contained within the text that has been read.

Other aspects of poetry make its linguistic features more salient or noticeable, including repetition of lexical items and syntactic structures, and also prosodic elements such as rhythm and rhyme. These last two have been cited as reasons why target items in poems have been found to be better recalled than those in prose (Tillmann & Dowling, 2007). Furthermore, drawing on a term applied to grammar learning but which might also apply to vocabulary, one might in turn anticipate that poetry prompts "learner-generated noticing" (Park, 2013, p. 75), whereby the reader notices the use of the linguistic feature themselves rather than having it being pointed out to them, potentially leading to stronger form-meaning connections and retention of the feature in question (Liu et al., 2021). Arguably that process may be similar to what is suggested in Share's (1995) self-teaching hypothesis which has mostly been applied to L1 learning but which has been extended more recently to L2 learning (Li & Wang, 2023).

It is also possible that teachers can help learners become more used to and proficient at noticing words in the input they encounter. That is one conclusion that can be drawn from the findings of a study with young beginner learners of French (Woore et al., 2018). Over six months learners read eight semi-authentic, challenging factual texts (rather than poems). Those whose teachers encouraged them to look carefully at language in the texts either through phonics or comprehension strategy instruction saw a much larger growth in general vocabulary size than learners who just read the texts and answered simple comprehension

questions, without any additional instruction. Given that there was no overlap between items in the test of vocabulary size (X-Lex, Meara, 1992) and vocabulary in the texts, the implication (although not stated by the authors of the study) is arguably that the effects of the instruction spilt over into how learners dealt with other L2 input they encountered. That leads us to consider the role of teaching approach in relation to different text-types.

Teaching approach: involvement

While attention and noticing are seen as cognitive aspects of engagement, engagement is also enhanced through emotional, more personal and experiential involvement (Philp & Duchesne, 2016) which may in turn aid vocabulary learning (Busse et al., 2020). Some text-types may facilitate emotional involvement more than others; thus poetry often focuses on emotions and thinking about abstract concepts (for example, love, loss, and so forth). Both theory and empirical studies related to the learning of abstract concepts suggest that the emotional valence of abstract vocabulary helps in its acquisition (Kousta et al., 2011). Furthermore, the prosodic aspects of poetry help heighten the emotional experience of the reader or listener, facilitating a simulation of "the emotions and state of mind that the poem is constructed to convey" (Johnson-Laird & Oatley, 2022: 5).

Factual texts can, of course, also include emotion-related vocabulary, although it is possible that by using figurative language rather than discussing emotions literally and directly, poetry can express ideas that in other forms would be ineffable or 'unspeakable' (Alexander, 2013) and hence treat challenging and possibly disturbing themes in a more sensitive way. Nevertheless, the type of response such emotion-related vocabulary elicits from learners and hence the level of engagement it prompts is in large part influenced by the teacher, through the pedagogical activities they select to draw attention to particular words or phrases. They might, for instance, direct learners to consider how words make them feel, what images, thoughts and emotions they conjure up, and how words are arranged for

stylistic effect as part of that process. By contrast, the teacher may decide to focus more on orthographic, phonological and syntactical aspects of the text, with a primary focus on form for its own sake and to aid literal and factual understanding. Such an approach, especially through tasks such as gap-fills and answering comprehension questions, has been identified as a common one for literary material in language textbooks in Russia (Calafato & Gudim, 2022) and can also be seen in teaching materials used in England (as in this example).

As outlined earlier, these two approaches, in relation to literary texts, can be termed a Personal Growth and a Language Approach respectively (Carter & Long, 1991). While the latter arguably elicits primarily cognitive involvement with language, the former also brings with it an affective or emotional dimension, and combining these different aspects of engagement may increase the depth in which language is processed (Phil & Duchesne, 2016). In other words, new information in a text is encoded with greater elaboration, a process which the Depth of Processing framework of Craig and Tulving (1975) indicates will lead to better learning and retention.

The Depth of Processing framework was, furthermore, the starting-point for an influential framework for instructed L2 vocabulary learning, the Involvement Load Hypothesis (ILH, Laufer & Hulstijn, 2001). According to the ILH, how much vocabulary is learnt incidentally depends on the degree of 'involvement load' induced by different activities that learners undertake while studying L2 input, typically written texts, without the explicit intention to learn vocabulary from it. Involvement Load is said to be influenced by three factors: need, search and evaluate. Need is conceived of as a motivational, noncognitive aspect of involvement, indicating the extent to which the learner really wants or needs to understand a word. A higher degree of need is believed to arise from self-generated 'need', that is, learners themselves decide or choose whether to try to understand a word or not, compared with need directed by the teacher or another external agent. For 'search', a higher

load is believed to be induced when the meaning or form of a word is looked up or retrieved from memory rather than when the meaning is provided (Laufer & Hulstijn, 2001). 'Evaluation' requires making judgements about words and their use, to think about their "form-meaning connection" (Yanagisawa & Read, 2010: 516), by, for example, considering how words compare with other words, comparing different word meanings in different contexts, or combining a word with others to create sentences (Hazrat & Read, 2022). While evaluation is referred to by Laufer and Hulstijn (2001) as a cognitive aspect of involvement, it could be argued that tasks requiring learners to make judgements about words might also introduce an emotional or personal dimension (for example, which word expresses a feeling most effectively). As such, the emotional element could have a "mediating" and potentially strengthening effect on cognitive involvement (Philps & Duchesne, 2016: 60), and hence promote better learning.

While a sizeable body of research has been conducted into the ILH, it has come under some scrutiny, because of inconsistent findings regarding its predictive power for the learning of target items. Indeed a recent meta-analysis concluded that the effect of involvement, while significant, is fairly limited, variable (perhaps because of some vagueness and inconsistency in how each component has been defined), and moderated by factors such as time on task and item repetition; different aspects of involvement also have different impacts on learning, with evaluation being especially important (Yanagisawa & Webb, 2022). Yanagisawa and Webb also commented, like Hazrat and Read (2022), on the lack of attention paid by research to the impact of 'need'. Furthermore, Laufer and Hulstijn (2001) themselves concluded that none of the various models of motivation they discussed in relation to the ILH shed light on how much need different kinds of learning activities bring with them. Other motivational frameworks such as self-determination theory (SDT, Ryan & Deci, 2020) might be more appropriate for explorations of 'need' (Yanagisawa & Webb, 2022), especially in contexts

like England where the present study took place, where instrumental reasons and motivation for language learning are persistently low and hence where fostering intrinsic motivation becomes vital (Lanvers & Graham, 2022). Within SDT, intrinsic motivation is driven by the fulfilment of three psychological needs, sense of competence, autonomy and relatedness (Ryan & Deci, 2020), respectively: feeling able to undertake a valued and challenging activity effectively; exercising agency and choice; and a sense of connectedness. Teachers who are 'autonomy supportive' value student perspectives and opinions, provide choice and open-ended activities (Ryan & Deci, 2020). Arguably, autonomy and relatedness seem the most relevant to 'need' in the ILH sense, in that they could be drivers for why learners focus on a word or not. They also link to the Personal Growth Approach for using literary texts in language classrooms, in so far as it advocates stimulating learners' engagement with the text by inviting them to relate it to their own experiences, feelings and imagination and to form their own personal responses to it. Such an approach could, however, also be possible with non-literary texts, again raising the question of how text-type might interact with teaching approach.

In summary, while there are theoretical arguments to suggest that exposure to literary texts, notably poetry, might boost vocabulary learning, the impact on linguistic aspects of L2 learning from studying such texts in class has been underexplored, especially among school-aged, relatively low proficiency learners. Still fewer studies have directly compared the impact of literary texts with that of non-literary texts. Whether such impact depends on the teaching approach employed with the texts has not, to our knowledge, been investigated either, or whether such an impact varies according to the type of vocabulary knowledge measured. For example, whether learning is limited to target items in the texts themselves, or whether it extends to more incidental learning of vocabulary that learners might encounter beyond the texts being studied. Theories of vocabulary learning also suggest, however, that

certain teaching approaches might be more beneficial than others, with potential advantages for those that maximise how far learners notice and process words with any depth by stimulating personal engagement with the text. The current study therefore considered the following Research Questions, explored among learners of French in secondary schools in England:

1. How much vocabulary is learnt through literary and factual texts respectively, in terms of learners' general vocabulary size? To what extent does the amount of vocabulary learnt depend on the teaching approach used with the texts?
2. How much vocabulary is learnt through literary and factual texts respectively, in terms of target items contained in the texts? To what extent does the amount of vocabulary learnt depend on the teaching approach used with the texts?

We hypothesised that learning gains would be greater from literary texts and from a teaching approach that stimulates personal engagement with texts studied. Our investigation into whether similar gains would be found for both target items contained within the texts and for general vocabulary size was more exploratory in nature. We did however anticipate that the poems would facilitate learning of target items in particular because of the increased noticing they might facilitate, while Personal Growth approaches might bring some triggering of 'learner-generated noticing' of other input encountered with potential benefits for more incidental learning that would show in general vocabulary size.

METHODOLOGY

Context

The study was part of a larger project that explored creativity and multilingualism (Authors xxx). One strand of that project considered creativity in the context of the teaching

of French at high school level in England, where levels of motivation for and proficiency in language learning are generally low (Ayres-Bennett & Carruthers, 2019). Rates of vocabulary learning and overall vocabulary size in particular have been characterised as low and lower than in other countries, with learners estimated to learn an average of between 2.7 and 1.73 words for every hour of instruction received (Milton, 2006; 2015). While the national curriculum implemented at the time of the study included the recommended study of 'literary texts in the language [such as stories, songs, poems and letters]" (DfE, 2013, np.), our observations as teacher educators suggested limited use of such texts by teachers, as Duncan and Paran (2017) also found in a survey of high school teachers in Europe. In order to establish an evidence-base that might inform classroom practice as well as make a contribution to knowledge, the present study explored the extent to which different teaching approaches and text-types might benefit classroom vocabulary learning. Regarding text-type, poems were selected as a *literary* type of text in which ambiguity, appeal to the emotions, unusual linguistic combinations are most marked and which would contrast most sharply with non-literary short articles, that is, *factual* texts. Regarding teaching approach, we considered two approaches that we called the *Creative* and the *Functional* approach respectively. The former was based on the Personal Growth Approach, the latter on the Language Approach (Carter & Long, 1991) outlined earlier (see *The intervention*).

Participants

From the larger study participant group (Authors xxx), complete sets of data for vocabulary learning (see *Vocabulary tests*) were available from 160 learners of French, aged 13-14, across eight schools from a range of geographical and social contexts across England. The majority of learners had been learning French at high school from the age of 11, having

had very variable amounts and forms of French teaching at primary school. On average learners were receiving between 1.5 and two hours of French instruction a week at school.

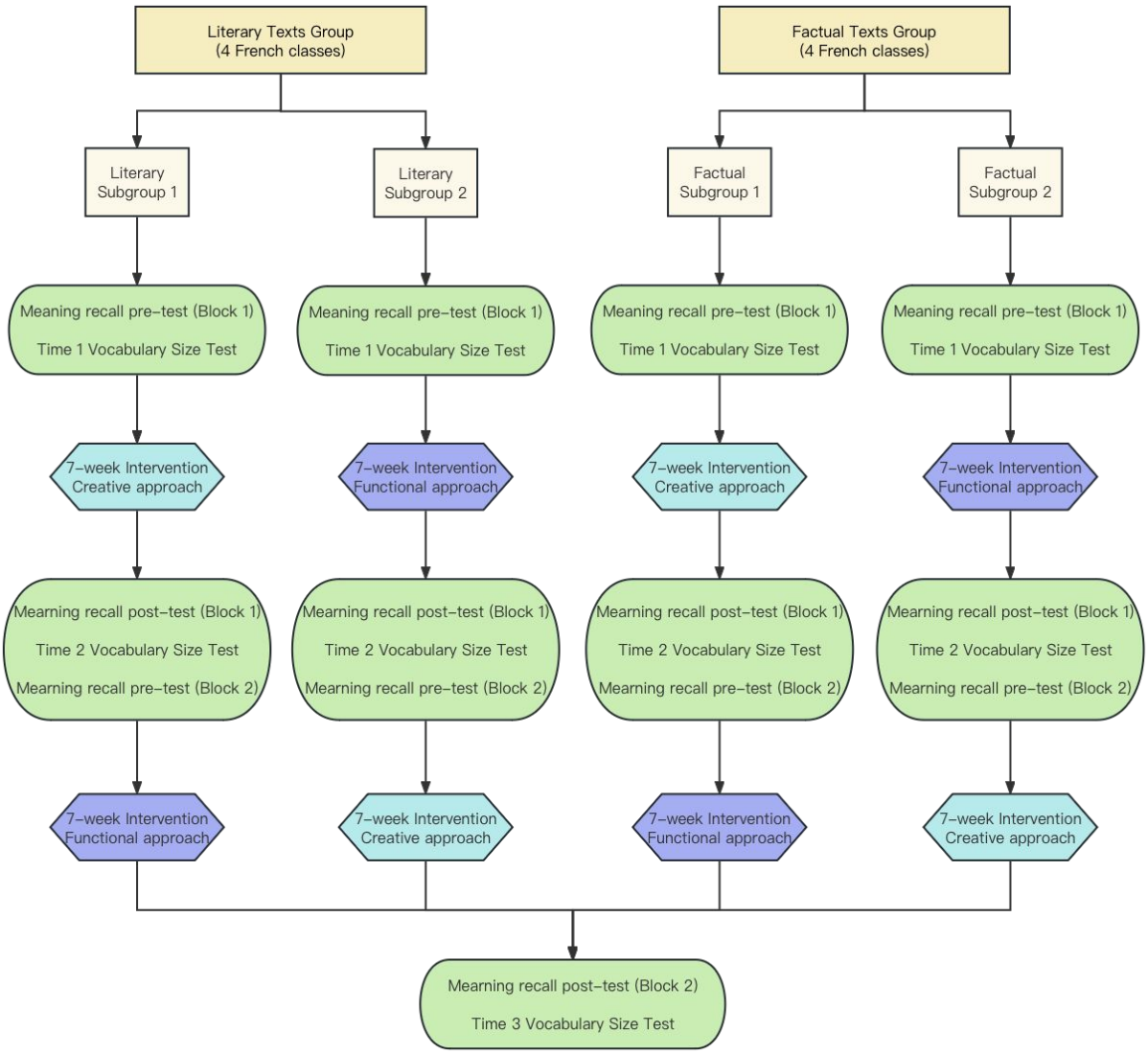
Schools were allocated to either a *literary* or a *factual* text group, with 81 learners in the former and 79 in the latter[1]. Efforts were made to achieve a balance of schools across the two groups in terms of: the socio-economic profiles of schools, as indicated by the percentages of pupils eligible for Free School Meals (FSM, a measure for social deprivation widely used in the UK); the number of pupils with English as an Additional Language (EAL); and number of pupils achieving a good pass in the GCSE, the examination taken at age 16 in England and hence an indication of levels of attainment in each school. Schools also provided data on the participants' prior academic attainment, which were standardised as z scores.

Study design

Both the *literary text* and *factual text* groups studied six texts in total, across two blocks of teaching (three texts per seven week block). In each block learners experienced either a *creative* or a *functional* teaching approach, following a counterbalanced design. That is, half of each text-group experienced the *creative* approach, and the other half the *functional* approach in the first teaching block, and vice versa in the second. Each teaching block began with a pre-test consisting of target vocabulary items that would be encountered over the teaching block. These same items were then reassessed at the end of the teaching block as a post-test. Learners' general vocabulary size was also assessed before and after the first teaching block, and for a final time at the end of the second teaching block (with the end of first teaching block test acting as a pre-test for the second teaching block). See Figure 1.

Figure 1

Study design



The intervention

As outlined in Figure 1, learners in both text groups experienced what we termed *creative* and *functional* teaching approaches. In the former, the focus was on generating learners' personal involvement with the text, and their attention was drawn to the emotional content. Any focus on linguistic features, such as use of tense, or vocabulary choice, was

undertaken with the aim of showing how they were used by authors to express meaning; learners were also asked to 'evaluate' such language use (Laufer & Hulstijn, 2001) in order to strengthen their level of personal and emotional engagement with the texts. For example, learners might be asked to judge which of two L1 translation versions of lines from an L2 text best conveyed the writer's feelings (see Figure 2 for further examples). Hence the *creative* approach drew on aspects of the Personal Growth Approach (Carter & Long, 1991), and sought to maximise learners' sense of connectedness through personal involvement with the text and autonomy in how they interpreted it. By contrast, in the *functional* approach, the aim was to generate learners' comprehension of information on a factual, literal and non-emotional level. Their attention was drawn to the grammatical, lexical and phonetic features of the text, and this focus on language patterns and linguistic features was undertaken with the aim of developing their grammatical, lexical and phonological knowledge. For both approaches we chose activity-types that schools in our sample would be familiar with using (based on our experience as teacher educators), for example, identifying examples of different grammatical tenses, gap-fill tasks, L1-L2 translation, matching words and pictures. As far as possible, the type of activity was kept constant across the two approaches for any given text but implemented in either a *creative* or a *functional* form. The exploitation of each text, regardless of approach, began with the glossing of words that learners would not be expected to know but that were necessary for understanding the text. Additionally, in both approaches, learners were given a small element of choice (in the activities offered for homework), but the *creative* approach alone gave learners scope for further autonomy, for example in open-ended in-class activities for which there was no right or wrong response and many different solutions were possible. Multimodal presentation was used in both approaches (video clips, images, sound) but in ways that corresponded to the focus of each (see Figure 2). To maximise fidelity to condition/approach, teachers received two hours of training by the

research team, materials (multimodal PowerPoint presentations) and lesson plans to use with the texts. These were designed so that 60 minutes of class time was spent on each text and an additional 20 minutes for out of class follow-up homework, to equalise time spent on activities across text-type and teaching approach (with approximately four hours in total spent on the texts in each teaching block). One observation of each teacher was also undertaken to monitor fidelity to condition.

Figure 2

The creative and functional approaches: example activities used with texts

Creative	Functional
Pre-reading: Background information about the author or context – emphasis on values, feelings.	**Pre-reading:** Background information about the author or context – emphasis on facts, dates, etc.
Look at images – prediction of emotions, themes addressed.	Look at images – prediction of events/information covered.
Key vocabulary given.	Key vocabulary given.
Working with the text: Listen to/watch YouTube rendition of the text, with accompanying music and written version of the poem.	**Working with the text:** Listen to the teacher reading the text aloud and follow the written version.
Verification/discussion of predictions	Verification/discussion of predications
L1-L2 translation – which of two versions is most expressive?	L1/L2 translation – which of two versions is the most accurate?
Look at verbs in future tense, negative expressions – what themes/emotions do they convey? Find examples and categorise them according to type of emotion.	Look at verbs in future tense and negative expressions. After a grammatical explanation, find examples of each feature and categorise them into type of verb and negative formation.
Listen again to the text and follow – how is emotion expressed through intonation?	Listen again to the text and follow – which phonemes are exemplified?
Post-reading: Summary – give your overall impression of text, what feelings it evoked	**Post-reading:** Complete summary gap fill- what can you remember about what happened in the text?

Texts

The poems used for the teaching intervention were carefully selected based on a range of relevant criteria. A crucial criterion regarding content was that each poem should treat a theme likely to be of interest to teenagers and capable of provoking some kind of reaction, but at the same time be accessible on a conceptual, emotional and linguistic level. The text length and complexity were important formal criteria. Prior to the intervention, the nine poems initially selected were shared with the teachers who would be involved in the study, as it was also important that they deemed the texts to be appropriate for their learners. They were asked to rate each poem for suitability of use, on a scale from 1 (not suitable at all) to 10 (highly suitable). This process resulted in five texts with a mean suitability score of over 6. An additional text was then selected in consultation with a teacher who had used it with learners of a similar age and proficiency level to those in the study.

These six poems were then analysed for readability, using the **Lix readability formula** which has been tested with a range of languages, including French **(**Björnsson, 1983**), LIX =A/B +(C/A x100)**, where

A = Number of words

B = Number of periods (defined by period, colon or capital first letter)

C = Number of long words (more than 6 letters)

Lix scores were then interpreted using the scale suggested by Björnsson (1983)**:**

20 = Very easy

30 = Easy

40 = Medium

50 = Difficult

60 = Very difficult

Following the selection of the target poems, we created or adapted factual texts to match the poems as closely as possible on the following criteria: readability level, topic, length and linguistic content. Thus, for example, one poem selected, 'L'homme qui te ressemble' [The man like you], deals with the theme of tolerance and common humanity. We wrote a short factual text on the issue of refugees seeking acceptance in Europe that used much of the same vocabulary and grammatical structures as the poem. Details of poems and matching factual texts, presented in the order in which they were used in each teaching block, appear in the Supplemental Materials, Appendix 1. The same themes were covered in each teaching block, and the readability level was balanced across the two blocks and across text types. Table 1 summarises features of texts and shows that all were within the Very Easy to Easy range (Björnsson, 1983). Independent-samples t-tests indicated that there were no significant differences across teaching blocks or text-types for readability, even though Block 2 texts were somewhat longer. It should be noted, however, that teachers in the study reported using these kinds of texts and themes very rarely under normal circumstances, and each of the texts would be considered challenging linguistically according to the expectations that are set for language learners in England (DfE, 2022).

Table 1

Summary of text characteristics

	Block 1		Block 2	
	Literary	**Factual**	**Literary**	**Factual**
	Mean (SD)		Mean (SD)	
No. of words	122.00 (16.70)	121.67 (14.10)	171.67 (105.16)	162.00 (53.56)
Lix formula	24.67 (9.29)	24.00 (6.00)	23.33 (8.51)	24.00 (1.00)

Vocabulary tests

Vocabulary knowledge was assessed in two forms: first, as general vocabulary size; second, as meaning recall for vocabulary items encountered in the intervention. These two forms were selected because our review of the background research suggested that reading literary texts might have an impact on learning of target items contained in the texts but also on vocabulary knowledge gained more incidentally beyond the target items. Both forms of tests were piloted before use.

Vocabulary size. Participants completed a version of the X-Lex French vocabulary test, adapted from the original X-Lex test (Meara, 1992) to make it suitable for lower-proficiency learners of French. We used three comparable versions of the test to avoid a practice effect, following Myles and Mitchell (http://www.flloc.soton.ac.uk/) from whose French Language Learner Corpora site the tests were drawn. The X-Lex test is a Yes/No test of vocabulary recognition, in this adapted version assessing 120 words: 20 randomly selected from each of the first five 1000-word frequency bands plus 20 non-words (included to allow

adjustments for guessing). Learners have to tick words they know, with a possible total score of 5000. We selected X-Lex for several reasons. First, word recognition is an aspect of vocabulary learning that typically occurs early on in the learning process (Nation, 2022). Our learners had relatively little lesson time for French overall (1.5-2 hours per week overall, see *Participants*) and spent only four hours approximately on the intervention in each teaching block. We therefore needed a measure that would capture smaller, more incremental change in vocabulary knowledge than could be achieved by assessing just learning of target items. Vocabulary size tests are also useful for assessing the impact of more incidental learning, which we anticipated might result from the intervention. Furthermore, there are very few, if any, well-evaluated tests of French vocabulary for lower-proficiency learners and a version of X-Lex has been used in similar studies (e.g. Graham et al., 2008; Woore et al., 2018) to assess changes in vocabulary size over around six months for school-based learners of French in England. X-Lex was also used by one of its developers, Milton (2006), with learners aged 11-18, indicating its suitability for the learners in our study. Using X-Lex also allowed us to make some comparisons with Milton's previous (2006, 2015) findings, as well as providing an easy to administer test (Uchihara & Clenton, 2020).

Meaning-recall vocabulary tests. These tests were designed to assess the learning of the meaning of certain target items encountered in the intervention. Each teaching block was preceded and followed by a vocabulary test containing 27 items (different items assessed for each teaching block). These items were given in written form in the L2 (French) and learners were asked to write the L1 (English) meaning.

Target items. In each text pair (*literary* and *factual*), key words/phrases were identified on the basis that they were important to the meaning of the text, were likely to be unknown to the learners, and yet from a relatively high frequency band. Thus, nine items per text were selected, giving 27 items that were included in tests for each teaching block. The

vast majority of such items occurred in the teaching materials that were used with the texts as well as in the texts themselves. For five target items occurring in Block 2, however, it was decided in retrospect that the teaching materials gave only minimal exposure to those target items, and that therefore they should be excluded from the subsequent analyses, giving 27 target items for Block 1, and 22 for Block 2. The Block 1 items for analysis comprised: 10 verbs/verb phrases; 10 nouns; four noun collocations; one adjective; one negative pronoun, one preposition. The Block 2 items for analysis consisted of: 11 verbs/verb phrases; 9 nouns; 2 prepositions/pronouns. In both Block 1 and 2 the majority of items (82% and 83% respectively) were from the K1 and K2 frequency bands. We selected items from different word classes in both Blocks to increase the ecological validity of the study, given that in a real language classroom context, words of different classes are taught and also encountered in authentic texts. Furthermore, as we explain in the *Data analysis* section, the statistical tests we used allowed us to control for any variability in difficulty level and so on across different items. All target items included in the analyses are given in the Supplemental Materials, Appendix 2.

Data analysis

The reliability of the X-Lex tests and the meaning-recall tests was measured using Revell's omega total, the greatest lower bound, and Coefficient *H* as alternatives for Cronbach's alpha (McNeish, 2018). Both instruments demonstrated good reliability (all measurements > .80, see Supplemental Materials, Appendix 3).

The data were then analysed using R (version 3.5.0; R Development Core Team, 2018) and the 'lmerTest' package (Kuznetsova et al., 2017). Linear mixed-effects models were employed for the analysis of the X-Lex tests because they had a continuous outcome variable. As the meaning-recall vocabulary tests had a binary outcome variable (correct/incorrect) they were analysed using generalised linear mixed-effects models. For

both types of test we had four fixed factors: 1. **Time** (Time 1 (Pre-test) vs. Time 2 (Post-test)); 2. **Text** (*Factual* vs. *Literary*); 3. **Approach** (*Creative* vs. *Functional*); 4. **Order** (the order in which learners experienced each teaching approach in the counterbalanced design, either Creative-Functional (CF) or Functional-Creative (FC)). For these categorical predictors, the following were set as the baseline for Time, Text and Order respectively: Time 1 (Pre-test); *factual*; *creative;* and CF. The fixed effects structure of the model was theoretically driven to answer the research questions, and hence included all four of these predictors and the highest four-way interaction between them. Learners' prior academic attainment data (*z*-scores) were included as a covariate to control for their existing level of academic performance in French.

We began analyses with a maximal random effects structure (Barr et al., 2013), that is, including all factors that might cause random variability in the dependent variable (either X-Lex scores or meaning-recall test scores), namely at the level of individual participant, the school they attended, and different test items (for the meaning-recall tests). The maximal model therefore included random intercepts for Participant and for School, by-Participant random slopes for Time, Approach, and their interaction, and by-School random slopes for Time, Approach, and their interaction. For the meaning-recall scores we also included random intercepts for Item, and by-Item random slopes for Time, Text, and Time x Text interactions. If the maximal model did not converge (that is, could not be computed), we gradually simplified it by removing interaction terms and the random effects which contributed to the least variance until a final converged model was obtained. The 'Anova' function from the 'car' package in R (Fox & Weisberg, 2019) was then used to establish whether fixed effects interactions were significant or not; those that were significant were then plotted using the 'effects' package (Fox & Hong, 2009) and further interpreted by conducting multiple pairwise comparisons with the 'emmeans' package (Lenth, 2019).

The analysis procedures outlined above allowed us to control for the fact that the intervention was delivered by different teachers (one teacher in each of eight schools), any differences in target item difficulty or frequency of occurrence in the texts (also controlled for by the counterbalanced design), and size of intervention group.

FINDINGS

RQ1. How much vocabulary is learnt through literary and factual texts respectively, in terms of learners' general vocabulary size? To what extent does the amount of vocabulary learnt depend on the teaching approach used with the texts?

For this Research Question the X-Lex test scores were analysed. Table 2 gives descriptive statistics by text-type, teaching approach, and test time points. Scores (especially at pre-test) were not dissimilar to those obtained by Milton (2006) for learners of similar age/proficiency level (including some learners scoring at zero).

Table 2

Descriptive statistics for the X-Lex tests (N = 160 in total)

Approach	Text	Time	N	Mean	SD	Min	Max
Creative	Factual	Pre-test	71	439.44	464.37	0	1850
	Factual	Post-test	71	728.39	489.32	0	1950
	Literary	Pre-test	77	653.90	449.62	0	1800
	Literary	Post-test	77	924.68	567.75	0	2450
Functional	Factual	Pre-test	67	620.90	428.36	0	2100
	Factual	Post-test	67	842.54	544.34	0	1850
	Literary	Pre-test	76	483.55	394.16	0	1600
	Literary	Post-test	76	824.34	465.58	0	1950

Note: The X-Lex test administered at the end of Block 1 was taken as the post-test for that block, and also as the pre-test for Block 2.

The data were then analysed using linear mixed-effects models. As the maximal model we started with did not converge, we simplified it to include only random intercepts for Participant and School, following the approach outlined in the *Analysis* section. Overall, the model showed a good-fit to the data. The two R^2 (Marginal R^2 = 0.21 and Conditional R^2 = 0.54) indicated that 21% of the variance was explained by the fixed effects and both fixed effects and random effects explained 54% of the variance.

Our analyses showed significant three-way Time x Text x Approach interactions ($X^2(1) = 5.83$, $p = .016$) and Time x Text x Order interactions ($X^2(1) = 4.03$, $p = .045$) (see Supplemental Materials, Appendix 4 for full model results). These significant interactions suggested that learning gains for each text-type not only differed according to which teaching

approach was used with them, but also according to which approach learners experienced first (remembering that all learners experienced both approaches, one each in each teaching block).

We first plotted the Time x Text x Approach interactions in order to interpret them (Figure 3) and then calculated vocabulary gains using multiple pairwise comparisons for "Time (Post-test vs. Pre-test)" by text-type and teaching approach (Table 3). Figure 3 suggests that there were learning gains for each text-type and teaching approach. Table 3 shows that such gains were all statistically significant. Furthermore, both the steeper line for the first and last plot of Figure 3, and the Estimates given in Table 3, indicate that the greatest gains occurred when *factual* texts were used with a *creative* approach, closely followed by the use of *literary* texts were used with the *functional* approach. There was an estimated vocabulary size increase of 354.58 and 341.07 respectively for these two text-approach combinations. Smaller gains were made by the *Creative* x *Literary* learners; using the *functional* approach with *factual* texts led to the smallest gains of all.

Figure 3

Time x Approach x Text effect plots

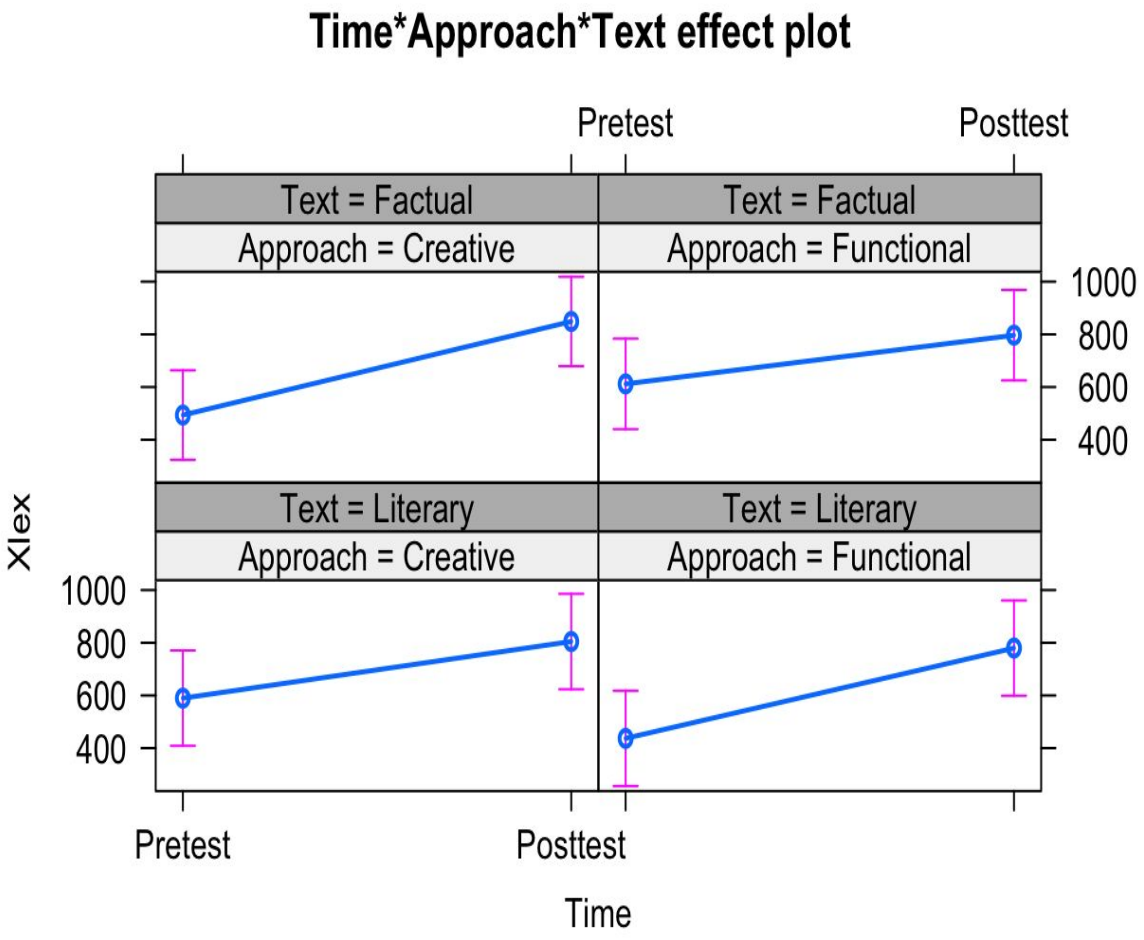


Table 3

Pairwise comparisons between test time points by text-type and teaching approach

Contrast	Text	Approach	Estimate	*95% CI*		*p*
				Lower	Upper	
Time2 - Time1	Factual	Creative	354.58	233.15	476.01	< .001
Time2 - Time1	Factual	Functional	188.28	63.37	313.19	.003
Time2 - Time1	Literary	Creative	206.57	85.89	327.26	< .001
Time2 - Time1	Literary	Functional	341.07	220.97	461.17	< .001

We next sought to understand the Time x Text x Order interactions. To do so we examined first the effect plots given in Figure 4, followed by multiple pairwise comparisons for "Time" by teaching approach and the order of the teaching approaches (Table 4). The first two plots in Figure 4 show the learning gains for the *factual* text group. The steeper line in the first plot suggests greater gains for learners whose teachers used the *creative* approach first (i.e. in Block 1), compared with learners who experienced the *functional* approach first (Plot 2). The estimates in Table 4 (297.53 vs. 245.33, row 1 and 2) support that conclusion. For the *literary* texts group, however, the opposite was the case. As is seen when the line slopes of Plot 3 and Plot 4 in Figure 4 are compared, there were larger gains for learners who experienced the *functional* approach first and the *creative* approach second compared with learners who started with the *creative* teaching approach. Again the estimates in Table 4 provide confirmatory evidence of that difference, which was large (371.80 vs. 175.85, row 3 and 4).

Figure 4

Time x order x text effect plots

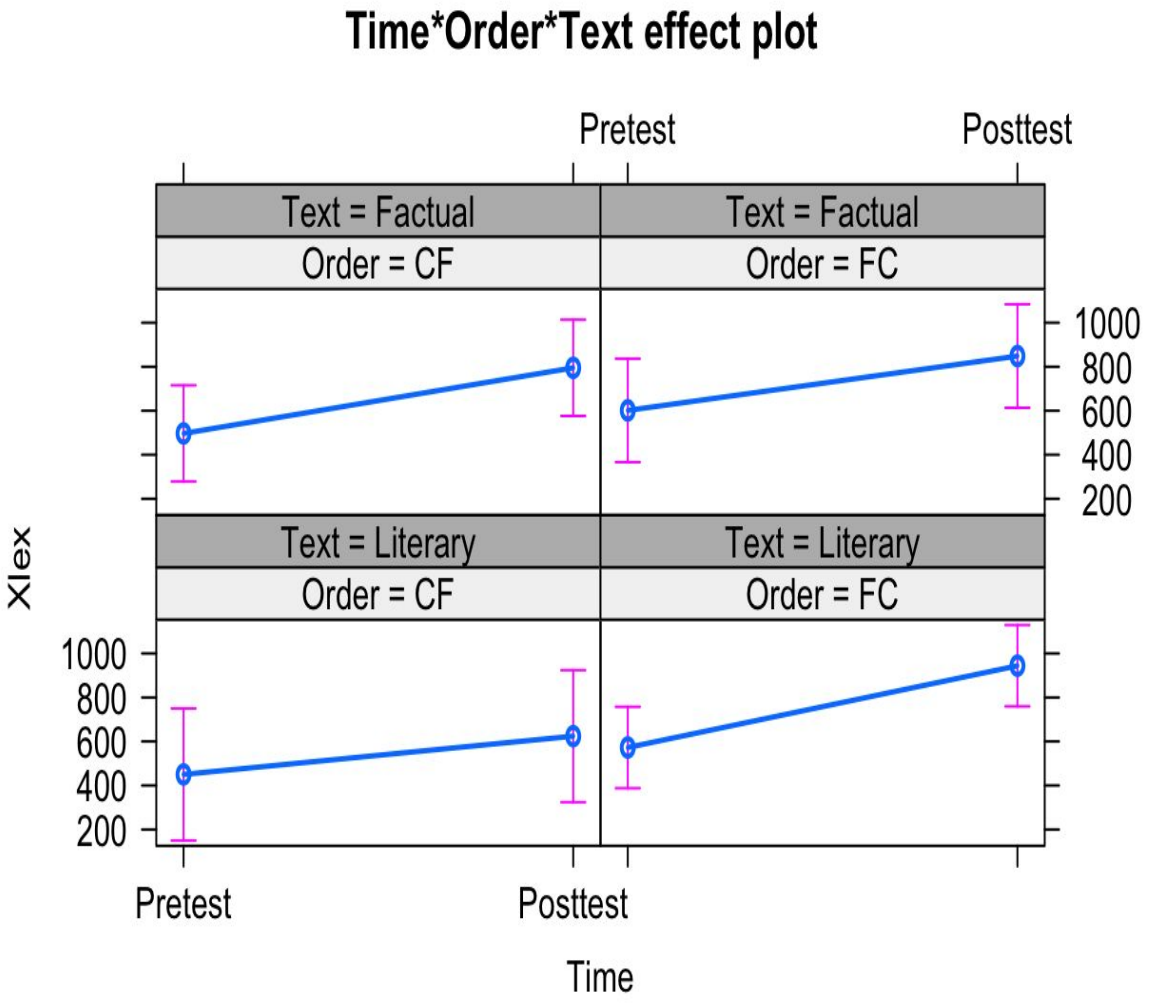


Table 4

Pairwise comparisons between test time points by text-type and order

Contrast	Text	Order	Estimate	95% CI		*p*
				Lower	Upper	
Time2 - Time1	Factual	CF	297.53	187.82	407.24	< .001
Time2 - Time1	Factual	FC	245.33	110.02	380.65	< .001
Time2 - Time1	Literary	CF	175.85	39.28	312.41	.012
Time2 - Time1	Literary	FC	371.80	270.11	473.48	< .001

RQ2. How much vocabulary is learnt through literary and factual texts respectively, in terms of target items contained in the texts? To what extent does the amount of vocabulary learnt depend on the teaching approach used with the texts?

To answer these questions, we first produced descriptive statistics for the meaning-recall vocabulary tests by approach, text-type, and test time point (Table 5). These showed that average learning gains were modest and in some cases low, but with quite a lot of variation across learners, as indicated by the Standard Deviations[2].

Table 5

Descriptive statistics for the meaning-recall tests (N = 160 in total)

Approach	**Text**	**Time**	**N**	**Mean**	**SD**	**Min**	**Max**
Creative	Factual	Pre-test	34	9.26	3.06	4.00	16.00

	Factual	Post-test	34	8.82	4.81	2.00	17.00
	Literary	Pre-test	60	6.82	2.73	2.00	14.00
	Literary	Post-test	60	11.32	4.45	4.00	25.00
Functional	Factual	Pre-test	62	5.32	2.46	1.00	12.00
	Factual	Post-test	62	7.56	4.31	1.00	25.00
	Literary	Pre-test	76	6.26	3.17	0.00	15.00
	Literary	Post-test	76	11.01	3.92	3.00	23.00

The data were then analysed using generalised linear mixed-effect models. Our final converged model included the random intercepts for Item, Participant, and School, by-Item random slopes for Time, Text, and Time x Text interactions, and by-Participant random slopes for Time. The two R^2 (Marginal $R^2 = 0.092$ and Conditional $R^2 = 0.67$) of the final model indicated that 9.2% of the variance was explained by the fixed effects; both fixed effects and random effects explained 65% of the variance. Full model results are given in the Supplemental Materials, Appendix 5.

As was the case for X-Lex, our analyses indicated that learning gains differed according to text-type, teaching approach used, and the order in which approaches were implemented. In other words, we found statistically significant three-way Time x Text x Order interactions ($X^2(1) = 22.57$, $p < .001$) and Time x Text x Approach interactions ($X^2(1) = 20.23$, $p < .001$).

In order to better understand the Time x Text x Approach interactions, the plots shown in Figure 5 were firstly produced. We then calculated gains using multiple pairwise

comparisons for 'Time' by text-type and teaching approach (Table 6). What is most striking in Figure 5 is the steep line for the *literary* text group, in both the third and the last plot, indicating the largest learning gains. By contrast, the line for the first, *Factual-Creative* plot slopes slightly downwards, indicating no learning. The *Factual-Functional* plot (Plot 2) lies somewhere between these two positions. These visual representations of learning gains are then supported in Table 6 by the Odds Ratios, which express the likelihood of learners scoring correctly at the post-test compared with scoring correctly at the pre-test. From those Odds Ratios we can conclude that the learners in the *literary* text group were around five times more likely to know the target items after the intervention, regardless of the teaching approach they experienced. By contrast, while the *factual* texts group did make statistically significant progress when they experienced a *functional* approach, they were only around three times as likely to know the target items after the intervention than before. Finally, when the *factua*l texts group experienced a *creative* teaching approach, they were no more likely to know the target items after the intervention than before, as is reflected in the Odds Ratio below 1 (Table 5, Row 1). Furthermore, although the effects of the *functional* approach seemed to be fairly comparable across the two types of texts (5.20 vs. 3.22), the benefits of the *creativ*e approach seemed to be much more strongly associated with *literary* texts use rather than with *factual* texts use (4.69 vs. 0.81).

Figure 5

Time x approach x text effect plots

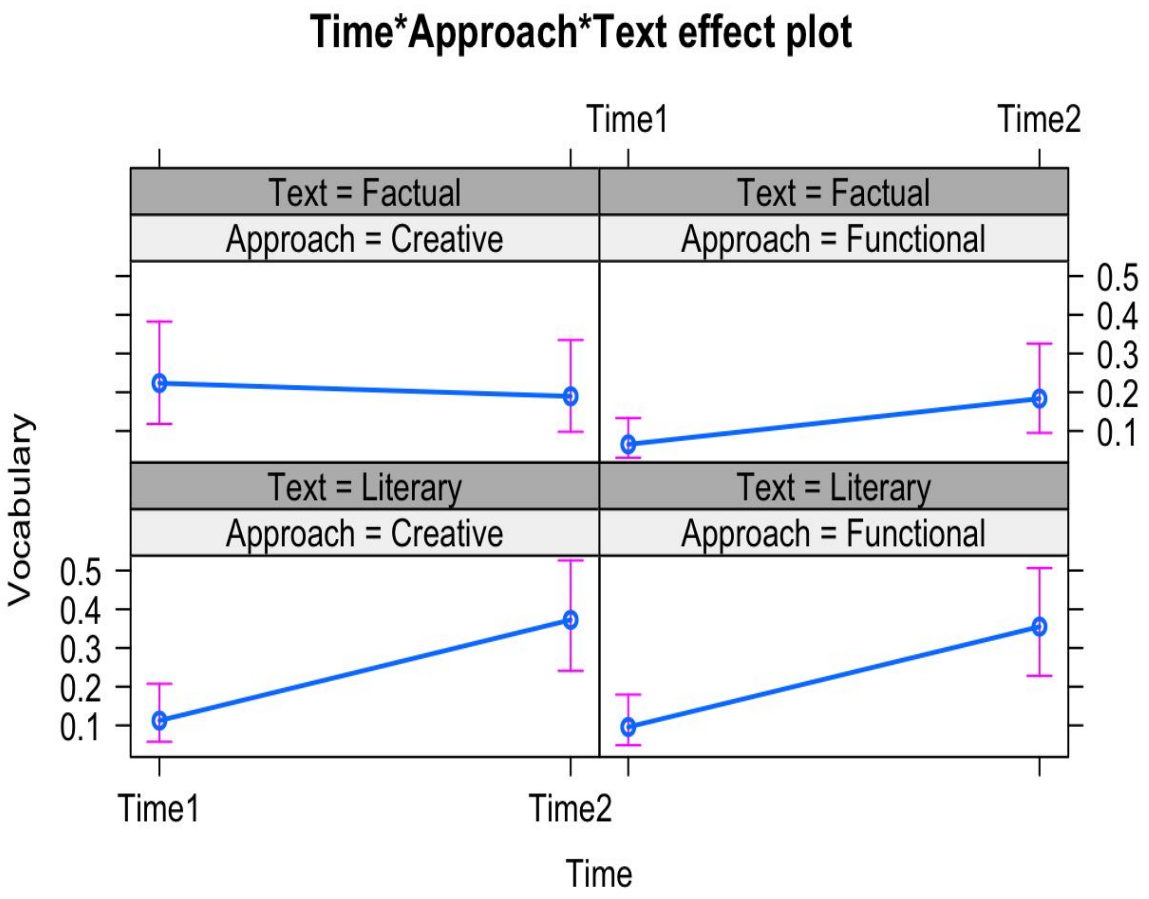


Table 6

Pairwise comparisons between test time points by text-type and teaching approach

Contrast	Text	Approach	Odds ratio	*95% CI*		*p*
				Lower	Upper	
Time2–Time1	Factual	Creative	0.81	0.52	1.26	.36
Time2–Time1	Factual	Functional	3.22	2.04	5.08	<.001
Time2–Time1	Literary	Creative	4.69	3.27	6.73	<.001
Time2–Time1	Literary	Functional	5.20	3.64	7.41	<.001

Finally, to better understand the Time x Text x Order interactions, we ran multiple comparisons for the effect of "Time" by text type and the order of the teaching approaches (see Table 7). The interactions are shown in Figure 6.

Figure 6

Time x order x text effect plots

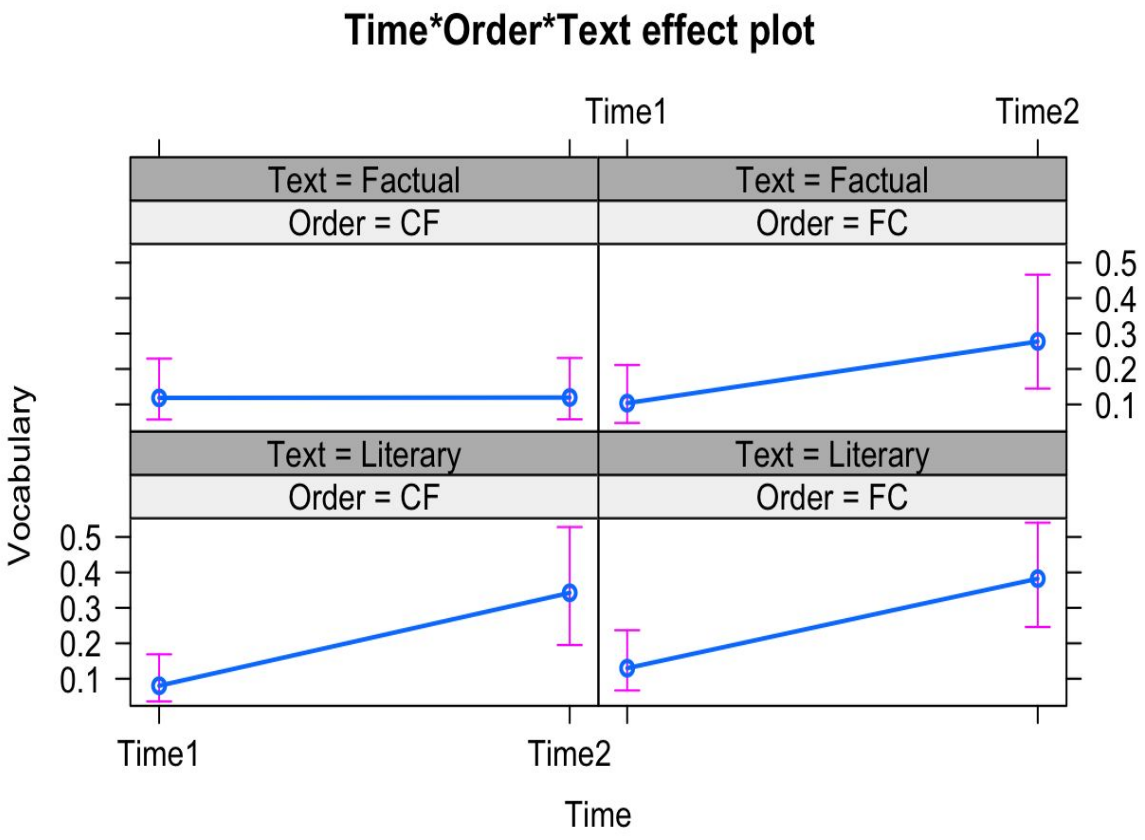


Table 7

Pairwise comparisons between test time points by text-type and order

Contrast	Text	Order	Odds ratio	95% CI		p
				Lower	Upper	
Time2–Time1	Factual	CF	0.85	0.55	1.32	.47
Time2–Time1	Factual	FC	3.07	1.91	4.96	<.001
Time2–Time1	Literary	CF	6.06	3.98	9.21	<.001
Time2–Time1	Literary	FC	4.02	2.81	5.75	<.001

These results firstly indicated that learners who read the *literary* texts made the largest, significant vocabulary gains, especially when they experienced a *creative* approach first and a *functional* approach second (Rows 3 and 4). For the *factual* text group, however, only those learners whose teachers started with a *functional* approach showed significant improvement in their knowledge of the target items. Experiencing first the *creative* approach, then the *functional* approach, did not lead to any significant vocabulary gains.

DISCUSSION

This study aimed to answer the following questions: How much vocabulary is learnt through literary and factual texts, in terms of a) learners' general vocabulary size; b) target items contained in the texts? To what extent does the amount of vocabulary learnt depend on how the texts are used?

Beginning with general vocabulary size, we found that learners, who had spent around four hours studying the texts in each teaching block, made gains in vocabulary size of between 188.28. and 354.58 words, depending on which text-type and teaching approach they experienced. That is rather remarkable, given the 2.7 to 1.73 words estimated to be gained on average for every hour of instruction received (Milton, 2006, 2015) by high school learners of

French in England below the age of 16. Direct comparisons should be made cautiously with those studies, however, given that their design was quite different from the present study (as was sample size, in the case of Milton, 2015). Furthermore, the average learning gains for target items contained in the texts in the present study were also much more modest, namely between two and five words per teaching block, but with variation across learners.

There was also variation in learning gains attributable to the text-type, teaching approach, and the order in which those approaches were experienced by learners. For general vocabulary size, the greatest improvement was made by learners who read *factual* texts and experience a *creative* approach, followed by learners who experienced the combination of *literary* texts and *functional* approach. A *Literary x Creative* combination led to slightly lower gains, and a *Factual x Functional* approach to the smallest gains. The order in which learners experienced each teaching approach also mattered: those who read *factual* texts did best when they experienced a *creative* approach before a *functional* one, but those who read *literary* texts benefited from having a *functional* before a *creative* approach.

Rather different conclusions can be drawn, however, regarding the learning of target vocabulary items contained within texts read. Learning gains were always greatest for reading *literary* texts rather than *factual* ones, and there was relatively little difference between using a *creative* or *functional* approach with the poems. The *functional* approach seemed comparably effective with either text-type, especially when it was used first and followed by the *creative* approach. By contrast, the *creative* approach only brought benefits (albeit strong ones) when used with *literary* texts, most strongly when used after a *functiona*l approach.

How might we interpret these somewhat complex and in places contradictory findings? First, different findings for general vocabulary size and learning of the target items in the texts themselves were not totally unexpected. The teaching intervention in all its forms

would be expected to have a more direct effect on learning of target items contained in the intervention texts, while learning gains for general vocabulary size suggest a more indirect effect, as the X-Lex test assessed knowledge of words beyond those included in the texts. That indirect effect on vocabulary knowledge may have arisen perhaps through some kind of triggering of learners' ability to learn from the wider French input that they were receiving. The *Factual x Creative* combination would have been the most novel one for learners, with its focus on the emotive quality of language in a factual text. That novelty may possibly have stimulated some kind of 'learner-generated noticing' (Park, 2013) or 'self-teaching' (Share, 1995) process. The fact that the smallest gains of all came from a *Factual-Functional* combination lends some support for such an interpretation. Also notable however is that for the greatest benefits learners needed to have experienced the *factual* texts in the *Creative-Functional* order, whereby their engagement was first stimulated in Block 1 through a *creative* approach before going on to a more *functional* consideration of language in Block 2.

Turning to the *literary* texts, the opposite was true, in that when order was taken into account, the largest learning gains for general vocabulary size occurred for the *Literary-Creative* (F-C) combination. That is to say, learners needed to have encountered what would have been unfamiliar and potentially more challenging poetic materials within a *functional* approach first, before going on to consider these in a *Literary-Creative* combination. Thus for poetic language to also stimulate the learner-generated noticing process referred to earlier, perhaps because of how such language "draws attention to itself" (Hanauer, 2001: 298), learners first needed to be able to access the poems on a more factual and straightforward level. Overall it seems that both *Literary* x *Functional* (F-C) and *Factual* x *Creative* (C-F) combinations may have achieved the best balance of challenge, engagement and accessibility.

Regarding the learning of the individual target items, where any effect of the intervention sessions would have been more direct, using *literary* texts consistently led to the

greatest gains, regardless of approach used. Individual words in the poems may have been more salient than in the *factual* texts, even within the same teaching approach, in line with Hanauer's (2001) view about poetry language cited earlier. In a questionnaire administered to learners at the end of each teaching block (not reported here for reasons of space) we also found that the *literary* texts were always enjoyed more, and significantly so for Block 2, although learners' prior academic achievement seemed to matter especially for the first set of texts in Block 1. Learners' greater enjoyment of the poems suggests that they perhaps experienced greater motivational 'need' to understand the words contained in those texts (Laufer & Hulstijn, 2001) and hence learnt them better than they did through the factual texts. Similarly, the fact that the *creative* approach seemed more effective when used with the *literary* rather than with the *factual* texts might further suggest that activities designed to heighten Personal Growth, 'evaluation, 'autonomy' and 'relatedness' work best for target item learning with texts that are in themselves more 'creative'. The *creativ*e approach worked much less well for target item learning with *factual* texts, especially if used first. This was the opposite of what was found for general vocabulary size, where a '*creative* approach first' was important for *factual* texts, in what we described earlier as a novel and unusual combination that seemed to have an indirect effect on general vocabulary size. As already argued, the target vocabulary items were perhaps less salient in the *factual* texts than in the poems, meaning that learners needed their attention drawn to them in a more language-focused way.

LIMITATIONS AND FUTURE STUDIES

Our study is limited by the absence of a control group that experienced the type of teaching and text normally used with learners aged 13-14 in England. We were able,

however, to compare the gains made in general vocabulary size by our learners with those from the same context in research by Milton (2006, 2015). Adding a control group would of course have further complicated our study design. Simplifying the latter, by, for example, just comparing text-type or teaching approach, might also have led to clearer conclusions. It would, however, have prevented us from exploring what seem to us to be a highly significant and previously unaddressed question, namely what kind of material used in what way leads to better learning outcomes. Future research could usefully explore that question with different languages and in different learning contexts. Finally, our interpretations of the present study's findings with reference to noticing are speculative only as we did not directly measure noticing or attention. Methods such as eye tracking would be useful in further studies to explore what learners attend to when reading different types of texts and in different teaching approaches.

CONCLUSIONS AND PEDAGOGICAL IMPLICATIONS

The current study took the original focus of investigating how much vocabulary is learnt from literary compared with factual texts and the relative benefits of different teaching approaches used with such materials. It established that learners of relatively low language proficiency are able to learn vocabulary from poems and semi-authentic, factual texts of a challenging nature. The relative benefits of different text-types and teaching approaches are complicated to determine, however. *Literary* texts emerged as more effective for learning target items; for general vocabulary size, *factua*l texts used creatively and *literary* texts used functionally both seemed to bring large benefits.

Taken together, these findings extend understanding of vocabulary learning in instructed settings, and especially through literary and other texts, in two important ways. First, they underscore the ways in which text-type and teaching approach interact and how that interaction may have differing impacts on different forms of vocabulary knowledge. Second, they suggest that there might be some merit in broadening out how 'need' within the Involvement Load Hypothesis is conceptualised to include aspects of Self-Determination Theory (Ryan & Deci, 2022), particularly autonomy and relatedness.

From a pedagogical perspective, there are a number of conclusions to be drawn from the study. Perhaps most significantly, for the England context at least, that lower proficiency learners are able to benefit in terms of vocabulary knowledge from studying texts that are of a more demanding nature, linguistically and cognitively, than the curriculum in England expects them to study (DfE, 2022). Proposals in England that take effect from 2024 and limit the range of language on which learners aged 16 will be assessed have been criticised for their potential washback effect on teacher practices, perhaps leading them to use only very simple and uninspiring texts in class, that give little if any exposure to lower frequency words (Milton, 2022). Given that in this study the benefits from reading more challenging texts were particularly strong for general vocabulary size, that prospect is a concern. The study also indicated that poems may be more useful than factual texts for the learning of specific items of vocabulary. Teachers should therefore be encouraged to experiment with their use and to raise their expectations for rates of vocabulary growth in their learners (Milton, 2015).

Finally, we found that for learning items of vocabulary it is generally better to start off with a *functional* approach and then move on to a *creative* one when using more challenging texts, factual or literary. By contrast, for increasing general vocabulary size, *factual-creative* and *literary-functional* combinations seem to work well, especially when learners first start working with such texts. In other words, no single approach can be recommended for any

particular text-type; instead, both have their respective merits and they should be used in conjunction with one another. That was the conclusion we shared with teachers involved in the project, and which has led to a set of recommendations for using the kind of texts we outline in this study, emphasising the value of using both *functional* and *creative* approaches. These recommendations, free to download here [link to be added post-review], include: the careful selection of texts that engage and challenge as well as covering key language; ways to stimulate learners' anticipation before they read the text in question; use of multimodality to support comprehension as well as to stimulate personal involvement; and different kinds of activities that encourage learners to engage with a text's language at a deeper level. Finally, while there is likely to be a degree of variation in terms of how much learners enjoy working with this kind of more challenging text, poems might be more engaging than factual texts.

In summary, while in many English as a Foreign Language classes the inclusion of literary texts is already established and especially for adult and more proficient learners, this is far from the case for the learning of other languages in input-poor settings such a England. Our findings suggest that high school language learning in such a context can also benefit from the use of literary texts alongside factual reading material, and that further research into their affordances is warranted.

Notes:

1. Gender is not reported as in the larger study a sizeable proportion of participants across both text groups opted not to report their gender, as was their right (we did ask about it but allowed a 'prefer not to say' option, a common practice in UK educational research). It is also not a focus of the present study.

2. It was not unexpected for the post-test SD to be larger than the pre-test SD as in classroom-based research interventions often do not benefit all learners equally. The greater variation at the post-test, was, furthermore, accounted for by including the by-Participant random slopes for test time points in the analysis.

REFERENCES

Authors xxx (to be added post-review)

Alexander, M. (2013). What use is poetry? *World Literature Today, 87,* 17–21. https://doi.org/10.7588/worllitetoda.87.5.0017

Ayres-Bennett, W., & Carruthers, J. (2019). *Policy briefing on modern languages educational policy in the UK. Belfast: MEITS/AHRC*. [Online] Available at: https://www.modernlanguagesleadershipfellow.com/ app/uploads/2019/01/Policy-Briefing-on-ModernLanguages-Educational-Policy-in-the-UK.pdf

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*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Björnsson, C. H. (1983). Readability of newspapers in 11 Languages. *Reading Research Quarterly*, *18*(4), 480–497. https://doi.org/10.2307/747382

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bloemert, J., Jansen, E., & van de Grift, W. (2016). Exploring EFL literature approaches in Dutch secondary education. *Language, Culture and Curriculum, 29,* 169-188, DOI: 10.1080/07908318.2015.1136324

Bobkina, J., & Dominguez, E. (2014). The use of literature and literary texts in the EFL classroom: Between consensus and controversy. *International Journal of Applied Linguistics & English Literature, 3,* 248- 259. https://doi.org10.7575/aiac.ijalel.v.3n.2p.248

Busse, V., Cenoz, J., Dalmann, N., & Rogge, F. (2020). Addressing linguistic diversity in the language classroom in a resource-oriented way: An intervention study with primary school children. *Language Learning, 70,* 82-419. https://doi.org/10.1111/lang.12382

Calafato, R., & Gudim, F. (2022). Literature in contemporary foreign language school textbooks in Russia: Content, approaches, and readability. *Language Teaching Research, 26*, 826-846. https://doi.org/10.1177/1362168820917909

Carter, R. A., & Long, M.N. (1991). *Teaching literature*. Longman.

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. https://doi.org/10.1037/0096-3445.104.3.268

Department for Education (DfE) (2013). *National curriculum in England: Languages programme of study*. https://www.gov.uk/government/publications/national-curriculum-in-england-languages-progammes-of-study/national-curriculum-in-england-languages-progammes-of-study

Department for Education (DfE) (2022). *French, German and Spanish GCSE subject content*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1076402/FGS_subject_content_09_05_2022.pdf

Duncan S., Paran A. (2017). The effectiveness of literature on acquisition of language skills and intercultural understanding in the high school context. The Hague: The International Baccalaureate Organisation.

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition, 42*, 499-523. doi:10.1017/S0272263119000494

Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software, 32*(1), 1–24. https://doi.org/10.18637/jss.v032.i01

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Graham, S., Richards, B., & Malvern, D. (2008). Progress in learning French vocabulary in a one-year advanced course at school. *Journal of French Language Studies, 18,* 349-364. doi:10.1017/S0959269508003499

Hanauer, D. I. (2001). The task of poetry reading and second language learning. *Applied Linguistics, 22,* 295–323. https://doi.org/10.1093/applin/22.3.295

Hazrat, M., & Read, J. (2022). Enhancing the Involvement Load Hypothesis as a tool for classroom vocabulary research. *TESOL Quarterly, 56,* 387-400. https://doi.org/10.1002/tesq.3051

Johnson-Laird, P.N., & Oatley, K. (2022). How poetry evokes emotions. *Acta Psychologica, 224,* 1-12. https://doi.org/10.1016/j.actpsy.2022.103506.

Kim, M. (2004). Literature discussions in adult L2 learning. *Language and Education, 18*, 145–166. http://dx.doi.org/10.1080/09500780408666872

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology, General, 140,*14–34. http://dx.doi.org/10.1037/a0021446

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

Lanvers, U., & Graham, S. (2022). Can we design language education policy and curricula for a motivated learner? Self Determination Theory and the UK language crisis. *The Language Learning Journal, 50,* 223-237. doi: https://dx.doi.org/10.1080/09571736.2022.2046353

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22,* 1-26. http://dx.doi.org/10.1093/applin/22.1.1

Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squared means*. https://CRAN.R-project.org/package=emmeans

Liu, Y.-T., Nassaji, H., & Tseng, W.-T. (2021). Effects of internal and external attentional manipulations and working memory on second language vocabulary learning. *Language Teaching Research*. https://doi.org/10.1177/13621688211030130

Li, Y., & Wang, M. (2023). A systematic review of orthographic learning via self-teaching. *Educational Psychologist, 58*, 35-56, DOI: 10.1080/00461520.2022.2137673

Maxim, H. H. (2002). A study into the feasibility and effects of reading extended authentic discourse in the beginning German language classroom. *The Modern Language Journal, 86,* 20-35. http://dx.doi.org/10.1111/1540-4781.00134

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Meara, P. (1992). *EFL vocabulary tests.* CALS University of Wales Swansea

Milton, J. 2006. Language lite: Learning French vocabulary in school. *Journal of French Language Studies* 16, 187–205. https://doi.org/10.1017/S0959269506002420

Milton, J. (2015). French lexis and formal exams in the British foreign language classroom. *Revue française de linguistique appliquée, XX,* 107-119. https://doi.org/10.3917/rfla.201.0107

Milton, J. (2022). Vocabulary denial and the false god of structuralism in Ofsted's 2021 Curriculum Research Review for languages. *The Language Learning Journal, 50,* 156-171, DOI: 10.1080/09571736.2022.2045680

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.) Cambridge University Press.

O'Donnell, M.E. (2009). Finding middle ground in second language reading: Pedagogic modifications that increase comprehensibility and vocabulary acquisition while preserving authentic text features. *The Modern Language Journal, 93,* 512-533. https://doi.org/10.1111/j.1540-4781.2009.00928.x

Paesani, K. W., Allen, H. W., Dupuy, B., Liskin-Gasparro, J. E., & Lacorte, M. E. (2015). *A multiliteracies framework for collegiate foreign language teaching*. Pearson.

Paran, A. (2008). The role of literature in instructed foreign language learning and teaching: An evidence-based survey. *Language Teaching, 41,* 465-496. http://dx.doi.org/10.1017/S026144480800520X

Park, E. S. (2013). Learner-generated noticing behavior by novice learners: Tracking the effects of learners L1 on their emerging L2. *Applied Linguistics, 34*, 74–98. http://dx.doi.org/10.1093/applin/ams016

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*, 97-130. https://doi.org/10.1017/S0272263115000224

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics, 36,* 50-72. doi:10.1017/S0267190515000094

Pouresmaeil, A., & Vali, M. (2023). The effects of incidental focus on form on learning vocabulary, grammar, and pronunciation. *Language Teaching Research*, *0*(0). https://doi.org/10.1177/13621688231185419

R Development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/J.CEDPSYCH.2020.101860

Schmidt R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.

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

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*, 329-363. https://doi.org/10.1177/1362168808089921

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Tillmann, B., & Dowling, W. (2007). Memory decreases for prose, but not for poetry. *Memory & Cognition, 35,* 628–639. https://doi.org/10.3758/BF03193301

Tsang, A., Paran, A., & Lau, W. W. F. (2023). The language and non-language benefits of literature in foreign language education: An exploratory study of learners' views. *Language Teaching Research*, *27*, 1120-1141. https://doi.org/10.1177/1362168820972345

Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, *24*, 540-556. https://doi.org/10.1177/1362168818799371

Woore, R., Graham, S., Porter, A., Courtney, L., & Savory, C. (2018). *Foreign Language Education: Unlocking Reading (FLEUR). A study into the teaching of reading to beginner learners of French in secondary school.* Final report to the Nuffield Foundation. Available at: https://ora.ox.ac.uk/objects/uuid:4b0cb239-72f0-49e4-8f32-3672625884f0

Yanagisawa, A., & Webb, S. (2022). To what extent does the Involvement Load Hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning, 71,* 487-536. https://doi.org/10.1111/lang.12444

Yang, A. (2001). Reading and the non-academic learner: A mystery solved. *System, 8,* 450–460. http://dx.doi.org/10.1016/S0346-251X(01)00038-0