Research Article

Samuel Barclay*

The effect of form presentation mode and language learning aptitude on the learning burden and decay of L2 vocabulary knowledge

https://doi.org/10.1515/iral-2023-0165
Received July 17, 2023; accepted July 3, 2024; published online July 23, 2024

Abstract: L2 words vary in how hard they are to learn and how quickly they are forgotten. This study explored two factors that may contribute to this variance: form presentation mode and language learning aptitude. English learners studied 32 words using electronic flashcard software in two conditions: unimodal (written form) and bimodal (spoken and written form). The frequency of exposure needed to learn each word was measured and used to determine learning burden. Learners, L2 users of English studying at a British university, completed immediate and two-week delayed tests, which were used to track any decay that occurred. Aptitude was measured using the LLAMA battery and incorporated into statistical modelling. The results showed that form presentation mode (bimodal > unimodal) and some aspects of aptitude impacted learning. No effects were found on decay. Pedagogical implications are discussed.

Keywords: vocabulary; learning burden; lexical decay; form presentation mode; language learning aptitude; flashcard software

1 Introduction

L2 vocabulary acquisition is influenced by various factors (Laufer 1997). These factors contribute to the learning burden of a word, making acquisition easier or more challenging for a learner. Although much is known about certain factors, there are others about which we currently know very little. Examples include form presentation mode and language learning aptitude. Word form is a crucial aspect of

***Corresponding author: Samuel Barclay**, Nottingham Trent University, Nottingham, UK,
E-mail: samuel.barclay@ntu.ac.uk. https://orcid.org/0000-0001-6112-382X

vocabulary knowledge that can be problematic for learners (Barcroft 2002, 2015). Word form can be presented via spoken and/or written modes, but to date limited research has investigated the comparative efficacy of these unimodal (i.e., written or spoken form) and bimodal (i.e., written and spoken form) conditions. Similarly, aptitude is a key individual difference (Dörnyei 2005; Williams and Burden 1997) that plays a crucial role in L2 acquisition (Li 2016). Research has considered its influence on certain aspects of L2 ability, but the relationship between aptitude and vocabulary learning has been relatively unexplored by studies to date. Certain factors can also impact the speed with which accrued L2 vocabulary knowledge decays (de Groot and Keijzer 2000). Lexical decay refers to the forgetting of L2 linguistic knowledge by healthy individuals in contexts of continued exposure to the target language (Barclay 2021; Barclay and Pellicer-Sanchez 2021). In vocabulary studies literature, limited research attention has been given to factors affecting decay, even though preventing such loss should, at least in part, shape most pedagogical decisions (Schmitt 2010). The current study investigated the impact of form presentation mode (i.e., bimodal v unimodal) and aptitude on the learning burden and subsequent decay of novel lexical items.

2 Background

2.1 Learning burden

The term learning burden describes the difficulty with which a word is acquired (Laufer 1997). Several factors combine to determine a word's burden, with models of burden typically including word, context, and learner-related variables (e.g., Higa 1965; Peters 2020). Understanding the contribution of these factors is important as it can help explain differential attainment and allow pedagogical stakeholders to build maximally efficient learning programmes. Studies have found that some factors lead to a heavier learning burden, slowing the learning process, while other factors ease the learning burden, leading to faster learning (see Peters 2020 for an overview). Learning burden is an amalgam of different factors that dynamically interact in complex ways. Crucially, therefore, burden does not exist in an absolute sense, but varies by-item and by-learner (Higa 1965).

Burden has generally been measured using a gains-based approach, which considers the number of target items learned during a fixed period of exposure, with more learning associated with less burden. However, some, although comparatively fewer, studies have employed a frequency-based metric, which considers the number of exposures needed for learning, with more exposures associated with a heavier burden (Barclay and Pellicer-Sanchez 2021; Higa 1965; Tinkham 1997; Waring

1997). Recent research has argued that a frequency-based approach is preferrable as it allows burden to be operationalised as a by-item, by-learner construct, whereas in gains-based approaches, responses are typically averaged across a group of learners or set of words (Barclay 2021). Furthermore, controlling exposure frequency, as is the case with a gains-based approach, may mean that some learners receive more exposures than they need to encode a target item and engage in retrieval. Studies indicate that retrieval frequency is linked to enhanced retention (Baddeley 2013; Karpicke and Roediger 2008; Royer 1973), suggesting that a gains-based approach may be inappropriate for an investigation that considers decay. The current study employed a frequency-based approach to investigate the contribution of presentation mode and aptitude to the learning burden of novel vocabulary.

2.2 Lexical decay

Research has shown that L2 lexical knowledge can be forgotten over a retention interval (Beaton et al. 1995; de Groot and Keijzer 2000; Ellis and Beaton 1993). Decay refers to the forgetting of accrued lexical knowledge despite continued contact with the target language. In experimental settings, loss is generally understood by comparing knowledge prior to (i.e., an immediate test) and following (i.e., a delayed test) a retention interval. In vocabulary research, although many studies have included delayed tests, research to date has generally focused on learning rather than loss; in fact, limited systematic consideration has been given to lexical decay to date.

The few studies that have considered this area suggest that, like burden, decay rates are impacted by several factors. Research has considered word, context, and learner variables, again finding that some factors are associated with decay, either speeding up or slowing down the rate of forgetting (e.g., cognateness - de Groot 2006; de Groot and Keijzer 2000; concreteness - Marefat and Rouhshad 2007; instructional method - Ellis and Beaton 1993; language distance - Hansen 2011; proficiency - Hansen 1999; Hansen et al. 2002; word length - Ellis and Beaton 1993; Tehan and Tolan 2007), while other factors seem less impactful (e.g., vocabulary size – Author). Interestingly, learning burden is associated with decay, with words that pose more burden more likely to be forgotten (Bahrick and Phelps 1987; de Groot 2006), speaking to the importance of considering learning burden in investigations of decay. Crucially however, few factors have received research consideration, and little is currently known about why some words seem to be forgotten faster than others, or why some learners appear to be more susceptible to loss. Two factors that could potentially impact the decay rate are the manner in which target items are presented to learners (Hill 1994) and learner aptitude (Bylund et al. 2009), the foci of the current study.

2.3 Form-presentation mode

Word form is a superordinate term consisting of spoken and written forms. The spoken form relates to the pronunciation of a word, while the written form relates to its spelling (Nation 2013). Independent presentation of the written or the spoken form is known as unimodal form presentation, while simultaneous presentation of the written and spoken forms constitutes bimodal form presentation. Within English language teaching (ELT), bimodal form presentation is generally preferred to unimodal presentation. It is recommended by teacher trainers (Scrivener 2011; Thornbury 2002; Ur 2010), and on teaching qualifications such as the Certificate in English Language Teaching to Adults (Thornbury and Watkins 2007); however, empirical research in this area is sparse.

Studies have shown that both unimodal written (Ebbinghaus 1885; Elgort 2011; Elgort and Piasecki 2014; Laufer and Shmueli 1997; Nakata 2015) and unimodal spoken (Bisson et al. 2015; Rodgers 1969) form presentation can facilitate L2 vocabulary learning. Although comparatively few studies have investigated bimodal form presentation, research suggests that it is also effective (Sandberg et al. 2011; Webb and Chang 2012). An obvious question to ask therefore, is which form presentation method is most beneficial for learning. Surprisingly, there have been few systematic comparisons of unimodal and bimodal form presentation methods in intentional learning contexts.

In one such study, Lado et al. (1967) compared unimodal and bimodal form presentation on the intentional learning of L2 Spanish by L1 English university students. Words were presented in various modality combinations (i.e., spoken only, written only, spoken before written, written before spoken, and simultaneous spoken and written), with 20 items presented in each condition. Results (meaning recall) showed the spoken only condition was least effective, the simultaneous spoken and written condition was most effective, and no difference was found between the other conditions. However, although the results supported the use of bimodal form presentation, there may have been a confounding item effect due to a lack of counterbalancing.

Hill (1994) also investigated this area, with L1 Cantonese learners of English studying target words in one of two conditions. All participants saw the written form, the phonetic transcription, an L1 equivalent, and an L2 definition with the experimental group additionally exposed to the spoken form of the target items. Results (spoken and written form recall) showed that the bimodal group had higher short-term and long-term gains in terms of spoken form knowledge. However, using the phonetic transcription in the written-only condition indirectly presented the spoken form to learners, blurring the lines between the two modalities.

Thus, although there is some evidence pointing to a positive effect of bimodal form presentation, this claim is based on limited evidence. Additionally, research to date has considered learning gains, while other metrics (i.e., learning burden) are perhaps better suited to evaluating the effect of different factors on the learning process. One study, Tinkham (1997), did consider the differential learning burdens of spoken and written unimodal form presentation modes, finding an advantage for written presentation; however, no comparison was made with bimodal form presentation and thus this question remains unanswered. Furthermore, although some studies have included delayed posttests (Hill 1994), no systematic consideration of the effect of presentation mode on lexical decay has been undertaken. Finally, studies to date have investigated form presentation mode in isolation, but as burden can vary from learner to learner (Higa 1965), any effect could be regulated by individual variables, speaking to the importance of considering the interaction between form presentation mode and potentially impactful learner variables. The current study met these gaps by using a flashcard learning platform to compare unimodal written and bimodal form presentation. The unimodal written condition was chosen as it is commonly used with paper flashcards and is often the default option on electronic flashcard platforms. The individual variable language learning aptitude was measured, and the interaction between aspects of aptitude and form presentation mode was modelled.

2.4 Language learning aptitude

Despite similar input, L2 learners differ both in the speed and the extent of linguistic mastery (Carrol and Sapon 1959). This asymmetry can partly be explained by language learning aptitude. Aptitude is "a specific talent for learning foreign languages which exhibits considerable variation between individual learners" (Dörnyei and Skehan 2003: 590). It relates to the time needed to complete learning tasks, with learners of higher aptitude needing less time to complete learning tasks. It also impacts the extent of acquisition that occurs given equitable learning and motivational conditions; meaning that comparatively able learners can acquire more language knowledge in a given timeframe (Carroll 1990). As it impacts the rate and extent of learning that occurs, it is an important variable in SLA (Skehan 1989).

Aptitude is generally understood as a multi-componential construct (Carroll and Sapon 1959; Granena 2014; Sternberg 2002); an amalgam of different cognitive and perceptual abilities. The specific cognitive variables included in an aptitude complex vary between constructs and measurement instruments, and are influenced by developments in cognitive psychology and differing test purposes. The

research presented in this paper considered associative memory capacity, i.e., the ability to associate a form with a label and retain this information (Carroll and Sapon 1959), sound recognition, i.e., the recognition of repeated spoken forms (Saito 2017), and phonetic coding ability, i.e., the capacity to distinguish specific sounds, to associate sounds with their written symbols, and to retain these associations (Carroll and Sapon 1959). The first relates to vocabulary learning (Granena and Long 2013) and the latter two aspects to spoken stimuli, which could potentially impact learning in bimodal presentation, speaking to a potential relationship between the target variables of this study.

For many years, aptitude research was marginalised as it was associated with structural approaches and considered less relevant to the communicative L2 classroom (Skehan 2002). In recent years, however, there has been a renaissance in this area (Wen et al. 2017), due, in part, to the availability of convenient measurement instruments (e.g., Meara 2005) and advances in cognitive psychology (Granena 2014). Research has shown that aptitude is related to language learning in L1 (Dörnyei 2005; Engel de Abreu and Gathercole 2012) and L2 (Ehrman and Oxford 1995; Granena 2014; Li 2016; Ortega 2009) and studies have found a connection between aptitude and specific aspects of L2 knowledge, with pronunciation (Granena and Long 2013; Saito 2019) and grammar (Granena 2014; Yalçın and Spada 2016) in particular receiving attention.

However, fewer studies have investigated the relationship between aptitude and L2 vocabulary acquisition. Granena and Long (2013) investigated the effect of age of onset, length of residence in the L2 context, and aptitude (vocabulary learning, grammatical inference, sound-symbol correspondence, and sound recognition) on the attainment of L2 phonology, vocabulary, and morphosyntax. The results indicated that, in general, aptitude (a composite of the different aspects) was strongly associated with the attainment of lexical knowledge in late-onset bilinguals ($r = 0.59$). In particular, phonological short-term memory and phonetic coding ability were strongly correlated with vocabulary learning. In his meta-analysis, Li (2016) found positive correlations between vocabulary learning and overall foreign language aptitude ($r = 0.15$). Additionally, he reported moderate correlations between vocabulary acquisition and associative memory capacity ($r = 0.20$) and phonetic coding ability ($r = 0.38$). Furthermore, Engel de Abreu and Gathercole (2012), investigating L2 acquisition with L1 Luxemburgish children aged seven to eight years old, found that phonological short-term memory (not considered in Li's meta-analysis) was also moderately associated with L2 vocabulary learning.

Thus, research has demonstrated that greater aptitude is associated with more successful L2 acquisition and various aspects of the aptitude complex are more or less strongly associated with the learning gains of different linguistic domains. Studies to date suggest that aptitude is positively correlated with vocabulary

acquisition; however, due to the paucity of studies linking aptitude with lexical acquisition, more evidence is needed. Motivation for the focus on aptitude in this paper stemmed from limitations not only in the quantity, but also the nature of research conducted to date. Investigations have generally looked at the relationship between vocabulary knowledge and aptitude, using this comparison to infer the effect of aptitude on the process of vocabulary learning (e.g., Granena and Long 2013). In such studies, a positive relationship between vocabulary knowledge and aptitude is typically used to argue that higher levels of aptitude led to faster learning. However, as correlations of attainment and aptitude may be impacted by confounding factors such as the amount of effort expended by a learner, it is important also to investigate the relationship between aptitude and the effort required to learn L2 lexical items. The current study addressed this gap by considering the relationship between components of aptitude and learning burden.

While some research attention has been given to the effects of LLA on vocabulary learning, it may be the case that LLA also impacts a learner's ability to retain accrued vocabulary knowledge. This points towards a potential relationship between LLA and language decay; however, research evidence in this respect is scarce. One of the few studies in this area investigated the attrition of L1 morphosyntactic knowledge (Bylund et al. 2009). A comparison of aptitude data with grammaticality judgement tasks indicated that aptitude was negatively associated with the maintenance of L1 morphosyntactic knowledge; attriters with above-average aptitude showed better retention of L1 structures than those with comparatively less aptitude. Based on these data, the researchers suggested that aptitude can offset the impact of reduced target-language exposure and argued that, given this effect, "it may be valuable to continue exploring the role of aptitude in attrition" (Bylund et al. 2009: 459). Although investigating implicit knowledge of L1, one of the goals of this paper is to investigate whether language aptitude plays a similarly preventative role in the decay of L2 lexical knowledge.

2.5 This study

Few investigations have considered the effect of form presentation mode and language learning aptitude on the learning burden and decay of L2 vocabulary knowledge. The current investigation addressed these issues by considering the following research questions:

1. To what extent is the learning burden of L2 vocabulary affected by form presentation mode and aspects of language learning aptitude?
2. To what extent is the decay of L2 vocabulary knowledge affected by form presentation mode and aspects of language learning aptitude?

Adult learners of English studied 32 lexical items using electronic flashcard software, with target word form presented in one of two conditions: unimodal written and bimodal. Learners interacted freely with the software, seeing each item as many times as necessary to correctly type the form twice. The frequency of exposure to each item was recorded and taken as the metric of learning burden. Target item knowledge was measured immediately after completion of the learning procedure and after a delay of two weeks. Decay was calculated by comparing target item knowledge on these assessments. Aspects of LLA were measured using the LLAMA (Meara 2005) battery and included in statistical models of learning burden (i.e., frequency of exposure) and decay.

3 Methodology

3.1 Participants

Participants (N = 58; 11 male, 47 female) were L1 Chinese (n = 28), Vietnamese (n = 21), and Thai (n = 9) students enrolled on an English for Academic Purposes course at a British university. Bilingual versions of the Vocabulary Size Test (Chinese: Nation *n*.d.; Vietnamese: Nguyen and Nation 2011; Thai: Nirattisai and Palanukulwong 2016) were administered, with results showing that learners had an average vocabulary size of 6,886 word families (SD = 1228.1). Although it is challenging to relate these scores to established proficiency levels (Wolter and Gyllstad 2013), Milton's (2010) findings suggest these learners had C1 proficiency, and thus were rather advanced. Learners also rated their proficiency in each of the four skills using a ten-point scale (one = extremely poor, almost no knowledge; ten = extremely good, almost native like). The mean scores were reading 6.47 (SD = 1.43); writing 5.24 (SD = 1.32); listening 6.50 (SD = 1.55); speaking 6.21 (SD = 1.45), again suggesting that the participants were somewhat advanced language learners. The participants had lived in an English-speaking environment for an average of 4.69 months (SD = 5.62 months), had studied English for an average of 11.75 years (SD = 5.67 years), and the average first contact with English was at 8.24 years old (SD = 3.12 years).

3.2 Learning software and target items

The target items from Barclay and Pellicer-Sanchez (2021) were used. Low frequency items were sampled from a frequency-based list compiled from the British National Corpus and the Corpus of Contemporary American English (Nation 2012). In total, 32

items were selected from the tenth and eleventh one-thousand-word bands controlling for referent familiarity, morphological transparency, and concreteness. They varied in length (number of letters) and part of speech (nouns and verbs). Prior to the study, two L1 Chinese, Vietnamese, and Thai speakers highly proficient in English confirmed that none of the target items were loanwords in the participants' L1s. Upon completion of the study, participants completed a checklist instrument to indicate prior knowledge of the target item and any exposure to an item during the study. Checked items were removed from analysis on a by-learner basis.

Due to the multilingual nature of the participants, L2 definitions were used to convey target item meaning. The definitions were graded so that the definitional language came from the most frequent 2,000 word families. Definitions were taken from Barclay and Pellicer-Sanchez (2021), where learners of lower proficiency than the current study reported that all definitions were comprehensible, and all definitional vocabulary was found to be unproblematic. Therefore it was highly probable that the more advanced participants knew the definitional vocabulary in the current study. To confirm this assumption, upon completion of the study, participants were asked to comment on the comprehensibility of each definition. No student reported difficulty. The target items and definitions are presented in Appendix 1.

The free flashcard platform *Anki* was used to present the target words, with items organised into decks of eight words. Studies have shown that electronic flashcards lead to sizable learning gains (Hung 2015; Nakata 2016) and are popular with learners (Hung 2015; Stroud 2014). Importantly, *Anki* allows items to be presented in both written and spoken form, facilitating the comparison of unimodal and bimodal conditions. Additionally, it provides data on the frequency of encounters that each participant has with each item. This metric was used to measure learning burden, with items that required more exposures to learn posing greater burden than items requiring fewer exposures.

In the unimodal condition, learners were presented with the written form only, while in the bimodal condition, the written form was followed by the spoken form after 400 ms. Initial piloting with ten participants of similar characteristics to the participants of the main study confirmed that learners were comfortable with this length of delay. The spoken forms were recorded by a professional voice actor and subsequently checked for intelligibility by an L1 speaker of English.

The flashcards were set up productively, with learners producing the target form to match a given meaning. After production, learners saw the correct form and were given feedback on the accuracy of their response. They then evaluated their production using three options: *Again, Good,* and *Easy*. *Again* was used when a learner had incorrectly produced or did not know a target form, *Good* when a learner had accurately produced a target form, and *Easy* when an item was known to a

learner prior to starting the experiment. The study employed a dropout procedure (Bahrick and Phelps 1987); when an item had been rated *Good* twice, it was removed from the item pool. In practice, a learner would be expected to select *Again* initially and then *Good* as their knowledge of a target item developed. The keystroke logging software InputLog (Leijten and Van Waes 2006), which records writing activity as text is composed on a computer (Spelman Miller and Sullivan 2006), was used to confirm the accuracy of these self-evaluations. A random sample of ten participants was used to confirm the validity of the self-report evaluation data, involving 3104 observations. For each learner, the key-stroke logging data was manually inspected and compared to the self-report evaluations. This showed that 99 % of evaluations were accurate, demonstrating that in the vast majority of cases, learners accurately evaluated their production.

3.3 Instruments

3.3.1 Immediate and delayed instruments

Tests of form recall and recognition were administered immediately after learning and once again after a two-week retention interval. Lexical decay was determined by comparing these tests, with items correctly answered on the immediate but not on the delayed test taken as having decayed. For form recall, learners saw the definition and typed the target form. Initial letter prompts were not provided, so this was a stricter measure of recall than some other studies (Laufer and Rozovski-Roitblat 2015) although minor spelling mistakes were accepted. The form recognition instrument consisted of multiple-choice items, with the key, three distractors, and an I don't know option to reduce guessing. Distractors were selected from the target item pool and an additional pool of non-target items matched with the target items for word class, length, and frequency.

Each instrument contained 40 items: the 32 target items and an additional eight novel items included to examine test effects. The test was adaptive; any item correctly produced on the form recall instrument was omitted from the form recognition test. This was to avoid giving positive feedback on learned knowledge that might have biased the delayed test. On the delayed administration of the instrument following a two-week interval, the order in which the target items were presented was varied, but otherwise the instruments were identical. Sample items are given in Appendix 2 and complete tests are available at: https://www.iris-database.org.

3.3.2 Language learning aptitude

Three tests from the LLAMA battery (Meara 2005) were employed: LLAMA_B, considered a measure of associative memory capacity; LLAMA_D, considered a measure of sound recognition; and LLAMA_E, considered a sound-symbol correspondence task measuring phonetic coding ability. These tests were chosen as their associated constructs were directly relevant to lexical acquisition and the foci of this study. LLAMA_B, for example, is considered to be a measurement of intentional vocabulary learning (Meara 2005), while LLAMA_D and LLAMA_E are thought to consider a learner's capacity to remember sound and associate it with written symbols, which may have impacted learning in the bimodal condition. For examples of second language acquisition research employing this test battery see Granena and Long (2013) and Saito (2017). Each test has a maximum score of one hundred.

Kolmogorov-Smirnov tests with the Lilliefors correction showed that the data were not normally distributed (LLAMA B: p = 0.04; LLAMA D: p = 0.05; LLAMA E: $p.$ = 0.01), and so it was necessary to split participants into groups. A median split was conducted, with participants equal to or above the median allocated to a high group and those below the median placed into a low group. Separate splits were performed for each aspect of aptitude.

- Associative memory capacity: Mean = 47.46 (SD = 21.12). Median = 45. High (range = 45–95) n = 33. Low (range 10–40) n = 25.
- Sound recognition: Mean = 28.66 (SD = 14.93). Median = 30. High (range = 30–60) n = 31. Low (range = 0–25) n = 27.
- Phonetic coding ability: Mean = 72.00 (SD = 22.17). Median = 70. High (range = 10–65) n = 36 | Low (range = 10–65) n = 22.

These values approximated the findings of Bokander and Bylund (2019), who administered the LLAMA instrument to 350 participants with different L1s, reporting the following mean scores: LLAMA B = 50.13; LLAMA D = 29.66; LLAMA E = 69.29. Thus, the values employed in the present study were largely in line with those of previous research involving learners from diverse contexts.

3.4 Procedure

The study was run in accordance with ethical approval and participants completed the approved consent form before taking part in the study. The study took place over

three sessions. In the first session, learners were inducted into the use of *Anki* and studied the 32 target items using the flashcard software in a computer room, with computers separated by partitions. Learners used headphones and the volume was checked prior to beginning the learning procedure. There was no time limit and learners worked at their own pace. The software was set up so that all words in a deck were presented either unimodally or bimodally. A within-participant counterbalanced design was employed. Half the participants learned the target words in a deck via unimodal form presentation and the other half through bimodal form presentation. Decks were ordered so that the form presentation condition alternated between decks. Additionally, the order in which the learners engaged with the decks varied by session. 32 learners were randomly assigned to the unimodal first group and 26 to the bimodal first group. Throughout the session, learners were monitored and could ask procedural questions.

Session Two took place the next day and consisted of studying the 32 target items and the immediate administration of the form recall and recognition tests. Completion of these instruments marked the onset of the retention interval.

Session Three took place two weeks later and began with administration of the delayed test. Once again, there was no time limit and participants worked at their own pace. Following completion of this instrument, a checklist instrument was conducted to determine prior knowledge of the target items and any intersessional item engagement. Next, learners completed the VST. Finally, an instructional LLAMA video was shown, and learners worked through the three tests. Each session lasted approximately 1 h. The procedure is summarised in Figure 1.

3.5 Data analysis

Prior to analysis, the data were trimmed using a checklist instruments of prior knowledge and intersessional exposure. Additionally, items which the frequency of

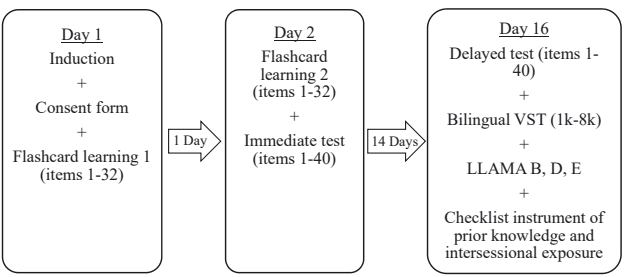

Figure 1: An illustration of the research design.

exposure data showed had been viewed only twice were removed. This procedure was adopted as two exposures equated to a learner selecting *Easy* during both Session One and Session Two, which they were instructed to do if they were familiar with an item. In total, 5 % of data points were removed across the cohort. Therefore, only novel items were included, as measured by the software and the checklist. Following this initial trimming, the frequency of exposure data were winsorised to mitigate the impact of outliers (Field et al. 2018).

The immediate and delayed tests contained eight items included to measure a test effect. These data were compared to determine the amount of learning that occurred from interaction with the target language on the immediate test. The results indicated that no learning occurred at the form recall level and minimal learning took place at the form recognition level (immediate test: m = 2.24 [SD = 1.84]; Delayed test: m = 3.04 [SD = 1.82]).

Data analysis was conducted in the statistical platform *R* (2016). Prior to statistical modelling, all continuous variables were log transformed. To determine the effect of the target variables on learning burden, a series of linear mixed effects models were computed using lme4 (Bates et al. 2015).

To investigate Research Question 1, separate models were calculated for frequency of exposure at the recall and recognition level. Importantly, only frequency data of items for which knowledge was demonstrated on the immediate test were included; for example, the form recall analysis only considered items for which knowledge was demonstrated at the level of form recall. The most parsimonious model was determined by a backward elimination procedure. The following regressors were fitted: form presentation mode, part of speech, word length, the VST score, L1, L3, familiarity with flashcards, group, LLAMA B, LLAMA D, LLAMA E. The inclusion of covariates was informed by previous research (de Groot 2006; de Groot and Keijzer 2000; Ellis and Beaton 1993; Lotto and de Groot 1998). Additionally, the interactions between form presentation mode and the three dimensions of aptitude were included in analysis. This facilitated consideration of the possible moderating role of aptitude on the efficacy of form presentation mode.

To answer Research Question 2, Generalised Logistic Mixed Effects Models were fitted to the delayed test data set. Logistic models were fitted because the outcome variable was binomial: a correct or an incorrect answer on the delayed test. Separate models were fitted to the recall and recognition data sets. Importantly, only those items shown to have been learned on the immediate test were included in the analysis. The best-fitting models were again determined by backward-elimination involving comparison of iterative models to establish the most parsimonious model. The following covariates were included in the maximal model: form presentation mode, PoS, word length, VST score, L1, L3, familiarity with flashcards, group, LLAMA B, LLAMA D, LLAMA E. Interactions between aptitude and form presentation mode

Table 1: Descriptive statistics (means with SD in parentheses) for the Immediate Test (Score) and frequency of exposure (Frequency) by mode of form presentation.

	Unimodal		Bimodal		Total	
	Score	Frequency	Score	Frequency	Score	Frequency
Form recall	12.09 (3.88)	8.37 (3.98)	11.97 (3.97)	8.08 (3.59)	24.05 (7.29)	8.23 (3.79)
Form recog.	14.66 (2.38)	8.86 (4.49)	14.05 (3.06)	8.54 (4.26)	28.71 (5.09)	8.69 (4.78)

were also included in analysis. Additionally, the learning burden (i.e., frequency of exposure during learning) was included in the decay analysis as previous research has shown a relationship between the difficulty of learning and the probability of decay (de Groot and Keijzer 2000).

4 Results

This section first presents results relevant to the impact of form presentation mode and aptitude on learning burden (Research Question 1) and subsequently considers the role of these variables in the process of lexical decay (Research Question 2).

4.1 Learning burden

Results of the form recall and recognition analysis are presented in turn. Descriptive statistics for the frequency of exposure needed to learn the target language are presented by condition in Table 1.

Descriptive statistics for the three aptitude tests were as follows: Associative memory capacity – M = 47.46 (SD = 21.12); Sounds recognition – M = 28.66 (SD = 14.93); Phonetic coding ability – M = 72.00 (SD = 22.17). Tables 2–4 display descriptive

Table 2: Descriptive statistics (means with SD in parentheses) for the Immediate Test (Score) and frequency of exposure (Freq.) by mode of form presentation and associative memory capacity.

			Unimodal		Bimodal		Total	
		n	Score	Freq.	Score	Freq.	Score	Freq.
Form recall	B High	33	13.00 (2.90)	7.54 (3.66)	12.82 (3.20)	7.26 (3.07)	21.72 (8.54)	7.40 (3.38)
	B Low	25	10.88 (4.67)	9.67 (4.12)	10.84 (4.64)	9.36 (3.96)	25.82 (5.70)	9.52 (4.04)
Form recog.	B High	33	14.73 (2.47)	7.81 (3.98)	14.36 (2.90)	7.68 (3.81)	29.09 (5.10)	7.75 (3.99)
	B Low	25	14.56 (2.31)	10.20 (4.74)	13.64 (3.26)	9.73 (4.55)	28.20 (5.14)	9.99 (4.65)

Table 3: Descriptive statistics (means with SD in parentheses) for the Immediate Test (Score) and frequency of exposure (Freq.) by mode of form presentation and sound recognition.

			Unimodal		Bimodal		Total	
		n	Score	Freq.	Score	Freq.	Score	Freq.
Form recall	D High	31	12.90 (3.21)	7.78 (3.35)	12.61 (3.57)	7.84 (3.29)	22.37 (8.31)	7.81 (3.32)
	D Low	27	11.15 (4.40)	9.15 (4.58)	11.22 (4.34)	8.39 (3.94)	25.52 (6.47)	8.77 (4.28)
Form recog.	D High	31	14.90 (2.23)	8.19 (3.83)	14.52 (3.05)	8.32 (4.07)	29.42 (4.91)	8.25 (4.37)
	D Low	27	14.37 (2.56)	9.65 (5.05)	13.52 (3.03)	8.81 (4.52)	27.89 (5.26)	9.24 (5.19)

Table 4: Descriptive statistics (means with SD in parentheses) for the Immediate Test (Score) and frequency of exposure (Freq.) by mode of form presentation and phonetic coding ability.

			Unimodal		Bimodal		Total	
		n	Score	Freq.	Score	Freq.	Score	Freq.
Form recall	E High	36	12.56 (3.55)	7.66 (3.25)	12.58 (3.37)	7.64 (3.19)	22.27 (8.31)	7.65 (3.22)
	E Low	22	11.32 (4.34)	9.65 (4.79)	10.95 (4.72)	8.91 (4.14)	25.14 (6.47)	9.28 (4.51)
Form recog.	E High	36	14.89 (2.15)	8.20 (4.00)	14.42 (2.88)	7.96 (3.64)	29.31 (4.86)	8.09 (4.25)
	E Low	22	14.27 (2.73)	9.96 (5.03)	13.45 (3.31)	9.54 (5.01)	27.73 (5.42)	9.76 (5.45)

statistics for learning burden by condition and the three aspects of language learning aptitude.

The best-fitting model for form recall is presented first followed by form recognition (see Appendix 3 for structures of all best-fitting models).

4.1.1 Form recall

Analysis of frequency of exposure showed a statistically significant effect for form presentation mode (t = -2.27, p = 0.02), indicating that items learned in the bimodal condition required fewer exposures that those learned in the unimodal condition. There were also statistically significant effects for associative memory capacity (t = -2.08, p = 0.04) and sound recognition (t = -2.02, p = 0.05), with the high groups needing fewer exposures to learn the target items than the low group. There was no effect for phonetic coding ability (t = -1.56, p = 0.12). A significant interaction between sound recognition and form presentation mode (t = 2.43, p = 0.02) was found, with participants in the low group benefiting from bimodal presentation more than the high group. Secondary effects were found for word length (longer items required

more exposures than shorter items), PoS (verbs needed more exposures than nouns), flashcard familiarity (more familiarity was associated with fewer exposures), L1 (Chinese learners needed more exposures than Vietnamese learners), and vocabulary size (learners with smaller vocabulary sizes needed more exposures).

4.1.2 Form recognition

Statistically significant effects were found for mode of form presentation (t = -2.68, p = 0.007) and associative memory capacity (t = -2.87, p = 0.006). The results indicated that the bimodal condition led to a lighter learning burden than the unimodal condition and learners with comparatively low associative memory capacity needed more exposures to learn the target items. No statistically significant effects were found for sound recognition (t = -1.05, p = 0.30) and phonetic coding ability (t = -1.04, p = 0.30). A statistically significant interaction was again found between sound recognition and form presentation mode (t = 2.42, p = 0.02), with the low group benefitting from bimodal presentation more than the high group. Secondary findings included an impact for word length (longer items required more exposures), and PoS (verbs required more exposures).

4.2 Lexical decay

Descriptive statistics for the decay of knowledge are presented in Table 5. This shows the mean number of items retained on the delayed test per condition. Additionally, the proportion of items retained relative to learning is shown. Importantly, analysis at each level only considered items for which knowledge was demonstrated at the equivalent level on the immediate test. For example, analysis of form recall data only included those items accurately produced on the immediate test. Overall, the table demonstrates that there was less decay at the level of form recognition than

Table 5: Descriptive statistics (means with SD in brackets) relevant to the decay of knowledge by condition.

	Unimodal		Bimodal		Total	
	M (SD)	Relative to learning	M (SD)	Relative to learning	M (SD)	Relative to learning
Form recall	5.14 (3.46)	0.40 (0.24)	5.29 (3.76)	0.42 (0.26)	10.43 (6.73)	0.40 (0.24)
Form recog.	13.02 (3.16)	0.90 (0.15)	13.27 (3.47)	0.93 (0.12)	26.29 (6.30)	0.92 (0.12)

Table 6: Descriptive statistics (means with SD in brackets) for the decay of knowledge by associative memory capacity.

			Unimodal		Bimodal		Total	
		n	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Form recall	B High	33	6.62 (3.73)	0.48 (0.25)	6.54 (4.01)	0.52 (0.28)	13.15 (7.21)	0.50 (0.25)
	B Low	25	3.87 (2.86)	0.30 (0.23)	4.20 (3.37)	0.28 (0.22)	8.07 (5.81)	0.29 (0.21)
Form recog.	B High	33	13.85 (3.86)	0.94 (0.24)	13.92 (4.03)	0.96 (0.24)	27.77 (7.71)	0.95 (0.24)
	B Low	25	12.77 (4.03)	0.85 (0.20)	12.23 (4.20)	0.89 (0.14)	25.00 (7.91)	0.87 (0.16)

Table 7: Descriptive statistics (means with SD in brackets) for the decay of knowledge by sound recognition.

			Unimodal		Bimodal		Total	
		n	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Form recall	D High	31	5.21 (3.68)	0.42 (0.23)	5.63 (4.25)	0.42 (0.28)	10.84 (7.38)	0.42 (0.24)
	D Low	27	5.11 (3.49)	0.38 (0.28)	5.11 (3.62)	0.40 (0.27)	10.22 (6.70)	0.39 (0.26)
Form recog.	D High	31	13.11 (2.56)	0.92 (0.19)	13.08 (4.45)	0.95 (0.19)	26.00 (5.71)	0.94 (0.18)
	D Low	27	13.35 (4.50)	0.88 (0.25)	12.89 (3.59)	0.90 (0.22)	26.43 (8.77)	0.89 (0.23)

form recall, with 60 % of accrued items forgotten at the recall level but only 8 % for form recognition. Furthermore, the table shows that there was considerable individual variation in the amount of loss that occurred.

The descriptive statistics for the three aspects of aptitude are presented in Tables 6–8. In general, it can be observed that in all cases the high groups outperformed the low groups with better retention relative to learning figures. However, there was considerable variance in the amount of decay that occurred.

4.2.1 Form recall

Results indicated that there was no effect for form presentation mode (z = 0.46, p = 0.65), showing that condition did not impact the decay of form recall knowledge. Similarly, no effects were found for associative memory capacity (z = 1.07, p = 0.29), sound recognition (z = 0.80, p = 0.43), or phonetic coding ability (z = -0.76, p = 0.45), and there were no significant interactions between mode and aptitude. Secondary effects

Table 8: Descriptive statistics (means with SD in brackets) for the decay of knowledge by phonetic coding ability.

			Unimodal		Bimodal		Total	
		n	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Form recall	E High	36	5.35 (3.51)	0.44 (0.24)	5.49 (3.82)	0.46 (0.26)	10.84 (6.84)	0.44 (0.23)
	E Low	22	3.00 (2.81)	0.34 (0.26)	3.20 (3.08)	0.34 (0.29)	6.20 (5.56)	0.34 (0.26)
Form recog.	E High	36	13.38 (3.96)	0.91 (0.22)	13.14 (4.25)	0.93 (0.19)	26.52 (8.01)	0.92 (0.20)
	E Low	22	12.33 (4.18)	0.90 (0.23)	12.00 (3.58)	0.92 (0.23)	24.33 (6.92)	0.91 (0.23)

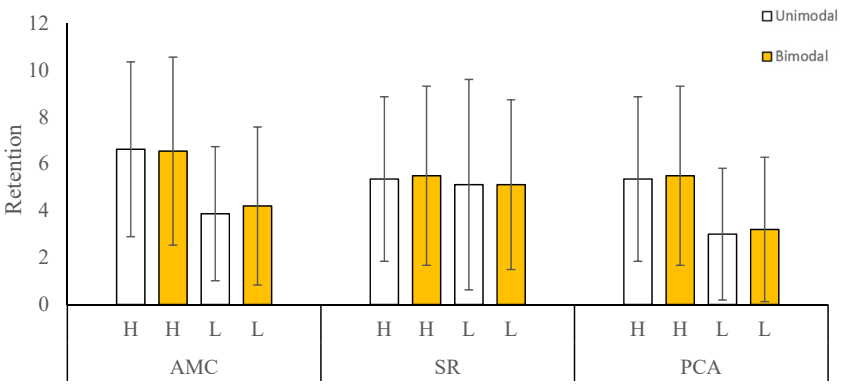

Figure 2: Descriptive statistics for the form recall delayed post-test by condition and aptitude. Note: AMC = associative memory capacity; SR = sound recognition; PCA = phonological coding ability; H = High; L = Low.

were found for learning burden, with frequency of exposure (z = -8.19, p < 0.001) significantly impacting decay: words that were more burdensome to learn were found to be more vulnerable to loss. Additional effects were found for PoS (knowledge of nouns showed less decay than knowledge of verbs) and L1 (Chinese L1 learners suffered significantly more decay than Vietnamese and Thai students). Figure 2 summarises the results for form recall.

4.2.2 Form recognition

Results for form recognition echoed form recall analysis. There was no effect for form presentation mode (z = 0.29, p = 0.77), indicating that lexical knowledge

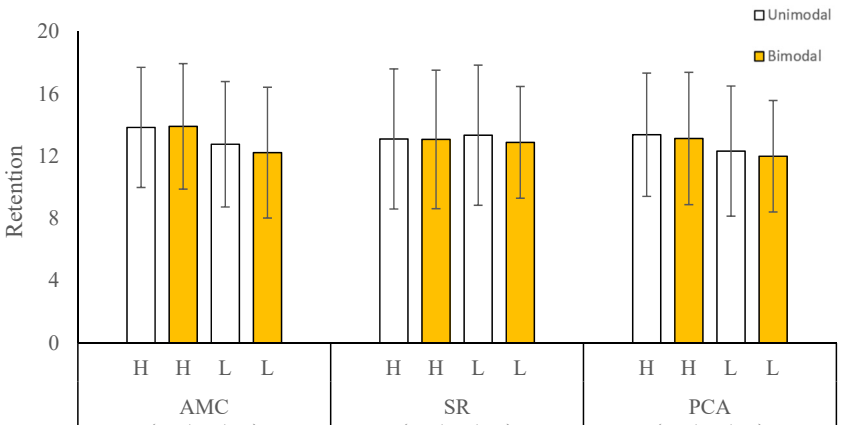

Figure 3: Descriptive statistics for the form recognition delayed post-test by condition and aptitude. Note: AMC = associative memory capacity; SR = sound recognition; PCA = phonological coding ability; H = High; L = Low.

acquired in the bimodal condition was as vulnerable as that accrued in the unimodal condition. Aptitude was also found to have no effect (associative memory capacity: $z = 1.85$, $p = 0.07$; sound recognition: $z = 0.28$, $p = 0.78$; phonetic coding ability: $z = -0.52$, $p = 0.60$). Furthermore, none of the interactions between mode and aptitude were significant. Secondary effects were again found for learning burden (frequency of exposure: $z = -2.03$, $p = 0.04$), with the words that posed more burden during the learning procedure found to be more vulnerable to loss. Additionally, significant effects were found for vocabulary size, with learners with smaller vocabularies more likely to suffer decay, and L1, with Chinese learners of English suffering more loss than the other two groups. Figure 3 illustrates the results for form recognition.

5 Discussion

In this section, general patterns are first summarised followed by a discussion of the findings related to each research question.

This study showed that the form-meaning connection of a large number of second language lexical items can be efficiently learned using flashcards, supporting other research in this area (see Nation 2013 for a summary). Learners required an average of eight exposures to learn the target items, supporting previous advice (Webb and Nation 2017). However, the results also support Huckin and Coady (1999) by showing the impossibility of establishing a frequency of exposure beyond which

learning is guaranteed. Some items required only three exposures to be learned while others needed thirty-nine. In general, the results indicated considerable individual and item variance. Given this variance, specifying a frequency of exposure at which learning occurs seems a Sisyphean task as any figure will inevitably overestimate or underestimate the challenge posed by certain items for certain learners.

With regard to the loss of knowledge, target item knowledge was found, unsurprisingly, to decay over the retention interval, supporting previous literature (de Groot and Keijzer 2000; Ellis and Beaton 1993). Greater loss was found at the level of written form recall than recognition, showing that form recall knowledge is more susceptible to decay than recognition knowledge.

RQ1: To what extent is the learning burden of L2 vocabulary affected by form presentation mode and aspects of the language learning aptitude?

The results showed that the learning burden was impacted by form presentation mode, with learners needing fewer exposures in the bimodal than the unimodal condition. Thus, form presentation mode did impact learning burden, with co-presentation of the written and spoken forms leading to less burden. This finding aligns with previous research; for example, Lado et al. (1967) also reported an advantage for bimodal form presentation. Hill (1994) however found no difference between unimodal and bimodal conditions, although in her study the phonetic transcription was presented to learners in the unimodal condition, allowing learners to indirectly access the spoken form of the target items. Therefore, the current findings are likely to better reflect the contribution of the form presentation variable to L2 vocabulary learning. In doing so, they provide evidence to support the advice common in practitioner discourse: teachers should endeavour to present the spoken and written forms of target items together.

Analysis showed that associative memory capacity (form recall and recognition) and sound recognition (form recall) significantly affected learning burden: learners with higher associative memory capacity and sound recognition needed fewer exposures to learn the target items. For associative memory capacity, participants in the low group required approximately 1.3 times the exposures to learn the target items than participants in the high group. For sound recognition, 1.12 times the exposures were needed. No effect for phonetic coding ability was found. Thus, it seems that to some extent aptitude plays a role in learning L2 vocabulary via flashcards. This finding partially supports the results of Granena and Long (2013), who report a positive relationship between aptitude and lexical attainment in late-onset bilinguals, and Li (2016), whose meta-analysis showed positive correlation

between associative memory capacity and lexical knowledge. Crucially, the current study expands this research. Previous studies have investigated the relationship between aptitude and attainment, while the current study considered aptitude and learning burden. In showing that learners with higher associative memory capacity and sound recognition learned L2 lexical items more quickly, this study suggests that this increased learning efficiency (and not for example, better retention) contributes to the differential attainment reported in previous studies.

Importantly, there was a statistically significant interaction between sound recognition and form presentation code, showing that learners with low sound recognition learned the target items more quickly in the bimodal condition than in the unimodal condition. For learners with higher levels of sound recognition, there was little difference. This is, to the best of my knowledge, the first study to have considered the interaction of language learning aptitude and contextual factors on the learning of L2 vocabulary, and more research is needed in this area.

RQ2: To what extent is the lexical decay of L2 vocabulary knowledge affected by form presentation mode and aspects of language learning aptitude?

No effect was found for form presentation mode on the decay of foreign language vocabulary knowledge, indicating that words learned in the unimodal and bimodal conditions were equally likely to be forgotten. Only one study to date has considered, albeit indirectly, the role of form presentation mode in the decay of vocabulary knowledge. Hill (1994) reported that there was less decay in a bimodal condition than a unimodal condition; however, the nature of the unimodal condition in that study differed from the present investigation, presenting the spoken form indirectly to learners via phonemic transcription. Therefore, the current study likely better reflects the effect of form presentation mode on lexical decay.

Bimodal presentation led to faster learning and did not have a detrimental effect on maintenance of learned knowledge. Therefore, materials writers and teachers should be encouraged to utilise this mode of form presentation. For teachers, this might include saying the target words aloud when writing them on the board, including the spoken form when developing electronic materials on educational platforms such as *Anki*, and encouraging learners to investigate the spoken form when searching for words in an electronic dictionary.

No effect was found for any aspect of aptitude on the decay process. Additionally, none of the interactions between aptitude and mode were significant. This means that although language learning aptitude impacted burden, it did not affect the loss of accrued knowledge over the two-week period. Stepping back, we might question why we would expect aptitude to impact the maintenance of lexical knowledge. Research

has shown aptitude to positively correlate with attainment (Li 2016). Attainment is a multivariate construct consisting of different memory processes included encoding, consolidation, and storage (Baddeley 2013). This study has shown that some aspects of the aptitude complex are involved in encoding memory traces; for example, learners with higher associative memory capacity encoded lexical items more quickly than others. This advantage would allow such learners to acquire more target items in a given period than less apt peers. Thus, research that reports positive correlations between aptitude and attainment might actually tap into differential levels of encoding. The current study shows that while encoding may be impacted by elements of the aptitude complex, consolidation and retrieval may not. To explain individual variance in decay patterns, it may therefore be necessary to consider the variables that impact consolidation and retrieval.

An alternative explanation for the nonsignificant effect of aptitude on decay relates to weaknesses in the LLAMA instrument. Despite extensive use in SLA research, a validity argument has yet to be fully realised for the LLAMA battery. In particular, the instrument as a whole and the various components have undergone minimal assessment of internal validity (Bokander and Byland 2019). More worryingly, recent studies that have begun to consider the validity of LLAMA B, D, and E suggest it has insufficient internal validity. For example, Bokander and Byland (2019) assessed the validity of LLAMA with 350 participants based in diverse contexts. Using classical item analysis, Rasch analysis, and principal component analysis, they found that LLAMA D had low internal consistency (α = 0.54), while LLAMA B (α = 0.81) and E (α = 0.74) were more acceptable. They also found that of the components, only LLAMA B had acceptable levels of discrimination. Additionally, they highlighted issues relating the construct validity of LLAMA D and E, suggesting that the former may tap into implicit language learning aptitude (also, see Granena 2014) and the latter is open to test strategies.

Replication using alternative instruments (e.g., MLAT, HiLAB) is needed to validate the current results. Furthermore, while the role of aptitude has received minimal attention in studies of vocabulary learning and thus the current study makes a useful contribution, the results need to be "treat[ed] with appropriate carefulness" (Bokander and Bylund 2019: 1). This is especially the case with regard to sound recognition and phonetic coding ability, the components of aptitude targeted by LLAMA D and E respectively. As Bokander and Bylund (2019) found that LLAMA B functioned sufficiently, conclusions in that regard can be less hedged: associative memory capacity was found to affect the learning process but not impact the process of decay.

6 Limitations

First, this investigation was only able to consider form-meaning aspects of word knowledge. While this is perhaps the most crucial aspect of vocabulary knowledge (Schmitt 2010) and reflects the most common use of flashcards (Webb and Nation 2017), there are clearly other dimensions involved in knowing a word (see Nation 2013). Future studies that investigate a broader spectrum of word knowledge aspects are needed. Furthermore, this study focused solely on knowledge of the written form. As such, it is likely to have missed some gains in spoken form knowledge, which would be particularly likely in the bimodal condition. This supports the case for bimodal form presentation; not only was it found to facilitate expedited acquisition of the written form but may have potentially led to increased gains in spoken form. Next, to encourage learners to participate, the learning phase of the study was prioritised in Sessions 1 and 2, and the aptitude battery was administered towards the end of Session 3. Ideally, cognitive tests should be conducted at the outset to avoid any interference from other research tools. Replication using such a design is therefore needed. Additionally, future studies might include duration of exposure alongside frequency as metrics of burden to provide a richer picture of the effort learners expended to learn the target items. Lastly, although this study showed that bimodal form presentation can expedite learning, the specific contribution of the spoken form to easing learning burden is unclear. In the bimodal condition, the spoken form was presented with the written form for each exposure. However, it might be that the presence of the spoken form is particularly beneficial during the initial exposures; alternatively, perhaps it is after the initial stages, when learners have partial knowledge of the written form that the spoken form supports learning. This is an empirical question and warrants future investigation.

7 Conclusions

The present study investigated the role form presentation mode played in the learning burden and decay of L2 vocabulary in an instructional context. It showed that when word form was presented simultaneously in the written and spoken modes, learners experienced less burden than when presented in the written mode only; however, there was no difference in the amount of decay suffered in the conditions. This suggests that teachers and learners should utilise bimodal form presentation where possible.

Additionally, some aspects of language learning aptitude were found to impact the learning burden of L2 vocabulary. This confirms what teachers have long witnessed in the classroom: certain learners acquire target items more quickly than others. This study has also shown that certain aspects of the aptitude complex interact with contextual factors. However, neither aptitude, nor any interactions with the learning activity, impacted the decay of target items. It may be the case though that other cognitive variables affect the decay process, speaking to the importance of further research in this area. Overall, this study has added to our understanding of learning burden and lexical decay by delineating the role of two variables, the effect of mode of form presentation and aptitude.

References

Baddeley, Alan. 2013. *Essentials of human memory*. New York: Psychology Press.

Bahrick, Harry & Elizabeth Phelps. 1987. Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13(2). 344–349.

Barclay, Samuel. 2021. Examining the learning burden and decay of second language vocabulary knowledge. London: UCL Institute of Education Doctoral thesis. Available at: https://discovery.ucl.ac.uk/id/eprint/10127853/1/Barclay_10127853_thesis.pdf.

Barclay, Samuel & Ana Pellicer-Sanchez. 2021. Exploring the learning burden and decay of foreign language vocabulary knowledge: The effect of part of speech and word length. *ITL – International Journal of Applied Linguistics* 172(2). 259–289.

Barcroft, Joe. 2002. Semantic and structural elaboration in L2 lexical acquisition. *Language Learning* 52(2). 323–363.

Barcroft, Joe. 2015. *Lexical input processing and vocabulary learning*. Amsterdam: John Benjamins.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 1(1). 1–48.

Beaton, Alan, Michael Gruneberg & Nick Ellis. 1995. Retention of foreign vocabulary learned using the keyword method: A ten-year follow-up. *Second Language Research* 11(2). 112–120.

Bisson, Marie-Josee, Walter van Heuven, Kathy Conklin & Richard Tunney. 2015. The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology* 68(7). 1306–1326.

Bokander, Lars & Emanuel Bylund. 2019. Probing the internal validity of the LLAMA language aptitude tests. *Language Learning* 70(1). 11–47.

Bylund, Emanuel, Niclas Abrahamsson & Kenneth Hyltenstam. 2009. The role of language aptitude in first language attrition: The case of pre-pubescent attriters. *Applied Linguistics* 31(3). 443–464.

Carroll, John B. 1990. Cognitive abilities in foreign language aptitude: Then and now. In Thomas S Parry & Charles W Stansfield (eds.), *Language in education: Theory and practice 74. Language aptitude reconsidered*. Virginia: Prentice Hall Regents/Prentice Hall.

Carroll, John & Stanley Sapon. 1959. *Modern Language aptitude test*. Psychological Corporation.

de Groot, Annette. 2006. Effects of stimulus characteristics and background music on foreign language learning and forgetting. *Language Learning* 56(3). 463–506.

de Groot, Annette & Rineke Keijzer. 2000. What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning* 50(1). 1–56.

Dörnyei, Zoltan. 2005. *The psychology of the language learner: Individual differences in second language acquisition*. New Jersey: Lawrence Erlbaum.

Dörnyei, Zoltan & Peter Skehan. 2003. Individual differences in second language learning. In Catherine J. Doughty & Michael Long (eds.), *The handbook of second language acquisition*. Oxford: Blackwell.

Ebbinghaus, Hermann. 1885. *Über das Gedchtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker and Humblot.

Ehrman, Madeline & Rebecca Oxford. 1995. Cognition plus: Correlates of language learning success. *The Modern Language Journal* 79(1). 67–89.

Elgort, Irina. 2011. Deliberate learning and vocabulary acquisition in a second language. *Language Learning Research Club* 61(2). 367–413.

Elgort, Irina & Anna Piasecki. 2014. The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2. *Bilingualism: Language and Cognition* 17(3). 572–588.

Ellis, Nick & Alan Beaton. 1993. Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning* 43(4). 559–617.

Engel De Abreu, Pascale & Susan Gathercole. 2012. Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology* 104. 974–986.

Field, Andy, Jeremy Miles & Zoe Field. 2018. *Discovering statistics using R*. London: Sage.

Granena, Gisela. 2014. Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics* 35(4). 483–503.

Granena, Gisela & Michael Long. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29(3). 311–343.

Hansen, Lynn. 1999. Not a total loss: The attrition of Japanese negation over three decades. In Lynn Hansen (ed.), *Second language attrition in Japanese contexts*, 142–153. Oxford: Oxford University Press.

Hansen, Lynn. 2011. The acquisition, attrition and relearning of mission vocabulary. In Monica S Schmid & Wander Lowie (eds.), *Modelling bilingualism: From structure to chaos*. Amsterdam: John Benjamins.

Hansen, Lynn, Yukako Umeda & Melanie McKinney. 2002. Savings in the relearning of second language vocabulary: The effects of time and proficiency. *Language Learning* 52(4). 653–678.

Higa, Masanori. 1965. The psycholinguistic concept of 'difficulty' and the teaching of foreign language vocabulary. *Language Learning* 15(3-4). 167–179.

Hill, Monica. 1994. A word in your ear: To what extent does hearing a new word help learners to remember it? In Norman Bird, Peter Falvey, Amy B. M. Tsui, Desmond M. Alison & Arthur McNeill (eds.), *Language and learning*. Hong Kong: Government Printer.

Huckin, Thomas & James Coady. 1999. Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 21. 181–193.

Hung, Hsiu-Ting. 2015. Intentional vocabulary learning using digital flashcards. *English Language Teaching* 8(10). 107–112.

Karpicke, Jeffrey & Henry Roediger, III. 2008. The critical importance of retrieval for learning. *Science* 319. 966–968.

Lado, Robert, Beverley Baldwin & Felix Lobo. 1967. *Massive vocabulary expansion in a foreign languagelanguage beyond the basic course: The effects of stimuli, timing, and order of presentation*. Washington: U.S. Department of Health, Education, and Welfare.

Laufer, Batia. 1997. What's in a word that makes it hard or easy? Intralexical factors affecting the difficult of vocabulary acquisition. In Norbert Schmitt & Michael McCarthy (eds.), *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.

Laufer, Batia & Bella Rozovski-Roitblat. 2015. Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research* 19. 687–711.

Laufer, Batia & Karen Shmueli. 1997. Memorizing new words: Does teaching have anything to do with it? *RELC Journal* 28(1). 89–107.

Leijten, Marielle & Luuk van Waes. 2006. Inputlog: New perspectives on the logging of on-line writing processes in a windows environment. In Kirk Sullivan & Eva Lindgren (eds.), *Computer keystroke logging: Methods and applications*. Amsterdam: Elsevier.

Li, Shaofeng. 2016. The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition* 38(4). 801–842.

Lotto, Lorella & Annette de Groot. 1998. Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning* 48(1). 31–69.

Marefat, Hamideh & Amir Rouhshad. 2007. Second language attrition: Are different nouns equally likely to be lost? *Porta Linguarum* 8(junio). 85–98.

Meara, Paul. 2005. *LLAMA language aptitude tests*. Lognostics. https://www.lognostics.co.uk/tools/LLAMA_3/.

Milton, James. 2010. The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In Inge Bartning, Maisa Martin & Ineke Vedde (eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. Amsterdam: Eurosla Monograph Series I.

Nakata, Tatsuya. 2015. Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition* 37. 677–711.

Nakata, Tatsuya. 2016. Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition* 39(4). 1–27.

Nation, Paul. 2012. The BNC/COCA word family lists. https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/information-on-the-BNC_COCA-word-family-lists.

Nation, Paul. 2013. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nguyen, Le Thi Cam & Paul Nation. 2011. A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal* 42(1). 86–99.

Nirattisai, Supika & Thanyapa Palanukulwong. 2016. A new measure for Thai students' vocabulary size: The English-Thai version of vocabulary size test. *Journal of English Studies* 14. 155–181.

Ortega, Lourdes. 2009. *Understanding second language acquisition*. Oxon: Hodder Education.

Peters, Elke. 2020. Factors affecting the learning of single-word items. In Stuart Webb (ed.), *The Routledge handbook of vocabulary studies*. New York: Routledge.

Rodgers, Theodore. 1969. On measuring vocabulary difficulty: An analysis of item variables in learning Russian-English vocabulary pairs. *IRAL* VII(4). 327–343.

Royer, James. 1973. Memory effects for test-like-events during acquisition of foreign language vocabulary. *Psychological Reports* 32. 195–198.

Saito, Kazuya. 2017. Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning* 67. 665–693.

Saito, Kazuya. 2019. The role of aptitude in second language segmental learning: The case of Japanese learners' English /ɹ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics* 40(1). 183–204.

Sandberg, Jacobijn, Marinus Maris & Kasper de Geus. 2011. Mobile English learning: An evidence-based study with fifth graders. *Computers & Education* 57(1). 1334–1347.
Schmitt, Norbert. 2010. *Researching vocabulary: A vocabulary research manual*. London: Palgrave Macmillan.
Scrivener, Jim. 2011. *Learning teaching: The essential guide to English language teaching*. Oxford: Macmillan Education.
Skehan, Peter. 1989. *Individual differences in second language learning*. London: Edward Arnold.
Skehan, Peter. 2002. Theorising and updating aptitude. In Peter Robinson (ed.), *Individual differences and instructed language learning*. Amsterdam: John Benjamins.
Spelman Miller, Kristyan & Kirk Sullivan. 2006. Keystroke logging: An introduction. In Kirk Sullivan & Eva Lindgren (eds.), *Computer key-stroke logging and writing: Methods and applications*. Leiden: Brill.
Sternberg, Robert. 2002. The theory of successful intelligence and its implication to language-aptitude testing. In Peter Robinson (ed.), *Individual differences and instructed language learning*. Amsterdam: John Benjamins.
Stroud, Robert. 2014. Student engagement in learning vocabulary with CALL. In Sake Jager, Linda Bradley, Ed Meima & Sylvie Thouësny (eds.), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference*. Dublin: Research-publishing.net.
Tehan, Gerald & Georgina Anne Tolan. 2007. Word length effects in long-term memory. *Journal of Memory and Language* 56(1). 35–48.
Thornbury, Scott. 2002. *How to teach vocabulary*. London: Longman.
Thornbury, Scott & Peter Watkins. 2007. *The CELTA course: Certificate in English language teaching to adults trainer's manual*. Cambridge: Cambridge University Press.
Tinkham, Thomas. 1997. The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research* 13(2). 138–163.
Ur, Penny. 2010. Teaching vocabulary. In Penny Ur (ed.), *A course in language teaching trainee book*. Cambridge University Press.
Waring, Robert. 1997. The negative effects of learning words in semantic sets: A replication. *System* 25(2). 261–274.
Webb, Stuart & Anna Chang. 2012. Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review* 68(3). 267–290.
Webb, Stuart & Paul Nation. 2017. *How vocabulary is learned*. Oxford: Oxford.
Wen, Zhisheng, Adriana Biedroń & Peter Skehan. 2017. Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching* 50(1). 1–31.
Williams, Marion & Robert Burden. 1997. *Psychology for language teachers*. Cambridge: Cambridge University Press.
Wolter, Brent & Henrik Gyllstad. 2013. Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition* 35(3). 451–482.
Yalçın, Şebnem & Nina Spada. 2016. Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in Second Language Acquisition* 38(2). 239–263.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/iral-2023-0165).