System 133 (2025) 103749

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

The roles of aptitude and working memory in L2 learners' processing and incidental learning of technical vocabulary during video-lecture-based tasks

Danni Shi [a,*], Andrea Révész [b], Ana Pellicer-Sánchez [b]

[a] Georgetown University, USA
[b] University College London, UK

ARTICLE INFO

Keywords:
Aptitude
Working memory
Technical vocabulary
Incidental learning
Lecture viewing
Eye-tracking

ABSTRACT

This study explored how aptitude and working memory affect second language (L2) learners' processing and incidental learning of technical vocabulary from a video-lecture-based task. Twenty-nine Chinese learners of L2 English performed the task, during which they watched an introductory lecture while taking notes. Eleven technical words presented in lecture diagrams were selected as target words. Participants' visual attention to the target words was captured using an eye-tracker. Following the task, participants completed an unannounced vocabulary post-test, a free recall test, and a set of aptitude tests. Two weeks after the treatment, a delayed vocabulary post-test and a battery of working memory tests were administered. Results from mixed-effects models revealed that working memory significantly predicted learners' learning of technical vocabulary. However, neither aptitude nor working memory emerged as significant predictors of learners' visual attention to the target words.

1. Introduction

Knowledge of technical words plays an important role in the academic success and professional communication of second language (L2) learners and users (Dang, 2020). With the increasing availability of multimedia resources, lectures, as a primary source of academic input in classrooms, provide learners with opportunities to learn new technical terms while simultaneously developing their understanding of subject content presented in multiple modalities. This process, known as incidental vocabulary learning, refers to learning of new words as a by-product of learners engaging in meaning-focused tasks, without being explicitly instructed to learn those words (Ellis, 1999). From a methodological perspective, it has also been operationalised as learning under conditions where learners are unaware of vocabulary post-tests following the task (Hulstijn, 2003). The present study adopted the methodological definition.

Despite the important role of lectures in facilitating L2 development, few studies have examined the simultaneous learning of technical vocabulary and subject knowledge from lecture viewing. Within the framework of task-based language teaching (TBLT), video-lecture-based tasks have emerged as a promising tool for such research. First, these tasks can be designed to incorporate auditory, textual, and pictorial information, enabling learners to encounter technical terms in rich contexts. Also, requiring learners to take notes during a video-based lecture helps them develop skills commonly used in real-life academic settings.

This article is part of a special issue entitled: TBLT in L2 Classroom published in System.

* Corresponding author. Georgetown University, 3700 O St NW, Washington, DC 20057, USA.
E-mail addresses: ds1938@georgetown.edu, dtnvdsh@ucl.ac.uk (D. Shi).

https://doi.org/10.1016/j.system.2025.103749
Received 30 December 2024; Received in revised form 9 June 2025; Accepted 11 June 2025
Available online 11 June 2025

It is apparent that video-lecture-based tasks, which involve simultaneously processing multimodal L2 input while engaging in note-taking, require a range of cognitive abilities. As such, performance on these tasks may be influenced by individual differences in aptitude and working memory, both of which are multi-component in nature and crucial for supporting various L2 learning processes (Baddeley, 2000; Skehan, 2002). So far, previous research on incidental vocabulary acquisition from L2 viewing has reported that sound recognition ability, an important component of aptitude, predicted meaning recognition (Muñoz et al., 2024), while complex working memory was significantly correlated with both form and meaning recognition (Montero Perez, 2020). However, little is known about how learners with different cognitive profiles approach video-lecture-based tasks. The findings of this research may provide valuable insights into the cognitive abilities involved in vocabulary acquisition from lecture viewing (e.g., effectively selecting, organising, and integrating multimodal L2 input), which could, in turn, inform the design of video-lecture-based tasks to help reduce cognitive demands and enhance learning gains in multimodal contexts.

To capture the multifaceted nature of the cognitive factors, we adopted a battery of aptitude and working memory tests. We also triangulated data from multiple sources, in the hope that by measuring learners' learning outcomes using vocabulary post-tests, capturing their real-time viewing behaviour through eye-movement recordings, and gaining insights into their conscious cognitive activities during viewing through their notes, a fuller picture of attentional and acquisitional processes during multimodal lecture viewing could be achieved. This approach aimed to help elucidate the cognitive components involved in real-time processing and those contributing to the learning of technical vocabulary from video-based lectures.

2. Literature review

2.1. Technical vocabulary and video-lecture-based tasks

Technical vocabulary generally refers to words used and known within a specific subject area, ranging from words that are used almost exclusively in a discipline and typically known only to experts (e.g., *allophone* in phonology) to high-frequency words that may or may not carry subject-specific meanings (e.g., *stop* and *tongue* in phonetics) (Liu & Lei, 2020; Nation, 2013). In the present study, we focus only on words that are used by experts in a subject field, meaning that learners are unfamiliar with both the new L2 forms and the concepts they represent. This is considered fundamentally different from much of incidental vocabulary acquisition that involves attaching new L2 forms to already known first language (L1) concepts and is more challenging than acquiring unfamiliar word forms alone (Schmitt, 2010). Therefore, the current study examines the simultaneous learning of both new subject knowledge and lexical forms, an area that has received little research attention.

Given the challenges associated with learning technical vocabulary, a combination of both explicit instruction and incidental exposure is recommended (Nation, 2013). While intentional learning may lead to better learning outcomes (Laufer, 2003), relying solely on this approach is inefficient, particularly due to the limited number of technical words that can be learned and the restricted exposure to those words in classroom settings (Webb & Nation, 2017). Therefore, incidental vocabulary acquisition through academic input serves as an important complement to the development of technical vocabulary knowledge. As a major source of academic input, lectures often provide rich contexts in multiple modalities (e.g., spoken explanations, written annotations, visual illustrations), allowing learners to infer the meanings of technical vocabulary and deepen their understanding of subject knowledge.

The theoretical support for this comes from Mayer's (2014) cognitive theory of multimedia learning, which has also served as the basis for much of the research on multimodal L2 reading (Pellicer-Sánchez et al., 2021). The model suggests that people indeed gain a deeper understanding of instructional content when being exposed to verbal and visual input simultaneously. As shown in Fig. 1, the model is based on the assumptions that there are separate processing channels for auditory/verbal and visual/non-verbal information, each with a limited processing capacity (Baddeley, 1992), and learning occurs when individuals actively engage in processing information by selecting relevant words and images, organising them into verbal or pictorial models, and integrating these with prior knowledge (Wittrock, 1989).

In addition to understanding how learners construct meaning from verbal and visual input, it is important to consider the cognitive demands imposed by note-taking, an integral component of academic lectures. On the one hand, note-taking may lead to dual-task interference (Pashler, 1994), as learners must allocate their attentional resources to both processing incoming information and recording lecture content (Piolat, 2007). On the other hand, note-taking may direct learners' overall attention to new information (Frase, 1970) and promote learning by encouraging active integration of this information with existing knowledge through paraphrasing, reorganising, and summarising. Such processes may facilitate deeper processing and more effective encoding of information into long-term memory (Craik & Lockhart, 1972). These perspectives align with the three main cognitive processes in Mayer's (2014)

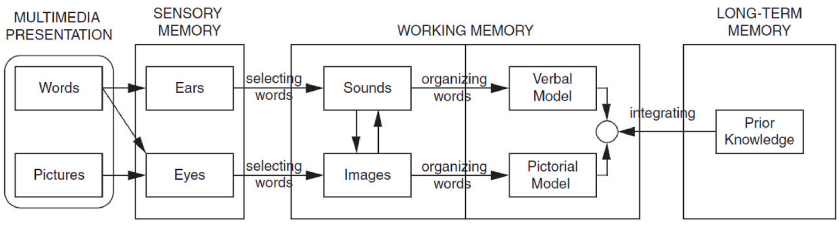

Fig. 1. Cognitive Theory of Multimedia Learning (Mayer, 2014, p. 52).

model, namely, selection, organisation, and integration. In addition, notes provide physical records of the lecture content from which information can be recalled and rehearsed during note-viewing (Di Vesta & Gray, 1973), further contributing to learning outcomes. Therefore, theoretical views on the role of note-taking in supporting learning outcomes remain inconclusive, with empirical studies also yielding mixed findings.

To explore how technical vocabulary is acquired through lectures when note-taking is allowed, TBLT appears to offer a promising pedagogical framework. Within the framework, viewing lectures while taking notes can be conceptualised as a video-lecture-based task that naturally meets all the key criteria of a task: a primary focus on meaning to achieve a clearly defined outcome (a set of notes), as well as opportunities for learners to use their own linguistic resources to process input and produce their notes (Ellis, 2003). Such tasks engage learners in cognitive processes similar to those involved in real-life academic situations (e.g., selecting, organising, and integrating multimodal information, determining what information to record, and monitoring whether incoming information aligns with previously recorded notes), allowing them to develop transferable cognitive and linguistic skills. Despite the pedagogical potential, research in this area is limited. While a few studies have examined the relationship between note-taking during auditory-only listening tasks and L2 comprehension (e.g., Chaudron et al., 1994), no research to date has explored the extent to which video-lecture-based tasks contribute to learners' processing and acquisition of technical vocabulary.

Investigation into the role of individual differences, particularly aptitude and working memory, might provide valuable insights, as these cognitive factors are especially relevant for understanding learners' task performance and outcomes (Li, 2024). Previous research on multimedia learning suggests that learners with high aptitude might be able to access "meaningful learning sets at will", thereby reducing their reliance on external support such as note-taking to activate these sets (Peper & Mayer, 1978, p. 521). Learners' ability to recognise sound patterns in oral input has also been reported to play a facilitative role in incidental vocabulary acquisition from L2 viewing (Muñoz et al., 2024). Furthermore, as discussed earlier, lecture comprehension and note-taking involve multiple operations, which require learners to strategically allocate cognitive resources due to the limited storing and processing capacity of working memory (Baddeley, 2000). Not surprisingly, working memory has been found to be significantly correlated with both participants' online reading behaviour and vocabulary gains (e.g., Huang et al., 2022). Therefore, the present study aimed to examine how different cognitive components underlying the constructs of aptitude and working memory affect the processing and learning of technical vocabulary through video-lecture-based tasks.

2.2. Aptitude and L2 vocabulary learning

Early aptitude research conceptualised language aptitude as a combination of cognitive and perceptual abilities that facilitate intentional and explicit L2 learning (Carroll, 1981). The classic model of aptitude put forward by Carroll and Sapon (1959) remains one of the most influential frameworks, which identifies four cognitive abilities essential for language learning: phonetic coding ability, grammatical sensitivity, inductive learning ability, and rote memory. Building on the early work, more recent research has begun to focus on how aptitude influences L2 learning in naturalistic settings. This line of research is particularly important, as learning that occurs in such contexts shares similarities with learning under contemporary teaching approaches that employ task-based methods (Long & Doughty, 2009). Accordingly, researchers have emphasised the need to assess learners' ability to acquire language not only explicitly (through reasoning and deliberate hypothesis testing) but also implicitly (through acquiring patterns in rich input without awareness of the rules) (Granena, 2013a). Given that incidental vocabulary learning may elicit both implicit and explicit learning processes, we measured the two types of aptitude in the present study.

Loosely based on Carroll and Sapon's (1959) model, Meara (2005) developed the LLAMA test, which has been widely used in SLA research and was adopted in the current study (for reservations regarding the validity of LLAMA as an aptitude measure, see Bokander & Bylund, 2020). The LLAMA test comprises four subtests, measuring rote, associative memory (LLAMA B), and the ability to recognise sound patterns in spoken language (LLAMA D), to associate sounds with symbols (LLAMA E), and to infer grammar of an unfamiliar language (LLAMA F). While LLAMA D has been claimed to have the potential to measure implicit learning aptitude (Granena, 2013b) or proceduralization (Suzuki, 2021), validation studies have yielded mixed results (Iizuka & Dekeyser, 2024), and additional research is needed to address the issue. Nonetheless, considering the distinct nature of LLAMA D, it may serve to investigate learning from multimodal input under non-instructed, meaning-focused conditions (Muñoz et al., 2024). Probably given that further validation of the LLAMA D subtest is warranted, the serial reaction time (SRT), a test adopted from cognitive psychology, has also been widely used for measuring implicit learning ability in SLA (e.g., Granena, 2013b; Kaufman et al., 2010; Linck et al., 2013).

To date, research has primarily focused on the relationship between aptitude and the level of vocabulary knowledge a learner achieves (e.g., Dahlen & Caldwell–Harris, 2013), but our understanding of how aptitude contributes to L2 vocabulary learning in naturalistic settings remains limited. Nagata et al. (1999) is one of the few studies within the TBLT framework that has examined the relationship between aptitude and vocabulary learning, although it exclusively focuses on the role of explicit aptitude. In their study, English-as-a-Foreign-Language (EFL) learners did an information gap task in which they followed directions to place kitchen objects in a matrix picture. Participants' aptitude was measured using sections of the MLAT and the Pimsleur language aptitude battery (PLAB). The results revealed significant correlations between participants' paired-associate and sound discrimination abilities and immediate meaning recognition, with paired-associate ability also significantly related to delayed meaning recognition.

Although not conducted within the framework of TBLT, studies investigating the role of aptitude in vocabulary acquisition from L2 viewing are also worth discussing. Suárez and Gesa (2019) and Teng (2022, 2024a) explored the effect of EFL learners' aptitude on vocabulary learning from captioned videos, with aptitude measured using the LLAMA total score (sum of the four subtest scores). Suárez and Gesa (2019) found that the total score was a significant predictor of meaning recall, while Teng's (2022) results showed that the total score was significantly associated with meaning and form recognition and recall. In contrast, Teng (2024a) reported that

aptitude did not significantly predict learning outcomes; instead, it had indirect effects on form and meaning recognition, mediated by L2 proficiency. This discrepancy might be due to the complexity and cognitive load of processing multimodal input in captioned videos, which could have diminished the direct impact of aptitude (Teng, 2024a).

While the use of the LLAMA total score prevented a clear understanding of the effects of individual cognitive abilities, Muñoz et al. (2024) examined the role of sound recognition ability alone, as measured by the LLAMA D subtest, in incidental vocabulary acquisition through repeated video viewing. Sound recognition ability was found to significantly predict participants' meaning recognition. To gain a more comprehensive understanding of the role of aptitude in vocabulary acquisition, additional cognitive abilities should be considered, such as associative memory, which is closely related to establishing form-meaning links.

Although these studies provide valuable insights into the potential influence of aptitude on vocabulary acquisition through L2 viewing, little is known about how aptitude affects learners' real-time language processing. Some processes (e.g., spoken word recognition) are considered highly automatic in nature (Field, 2013), making it difficult to investigate using verbal reports. To address this limitation, eye-tracking has been widely used in L2 reading and viewing research, as it provides a window into visual attention by capturing a wide range of viewing behaviour, while allowing learners to engage with reading or viewing without posing a secondary task (Conklin et al., 2018). Despite its methodological advantage, few studies have adopted eye-tracking to examine how aptitude affects online lexical processing during viewing. Hence, the current study adopted eye-tracking to address this gap. The insights gained from the current study could contribute to our understanding of the cognitive abilities involved in vocabulary learning through lecture viewing, thereby informing language teaching practices that support learners with different cognitive profiles.

2.3. *Working memory and L2 vocabulary learning*

In addition to aptitude, working memory has emerged as another important cognitive factor influencing L2 vocabulary acquisition. It is broadly defined as "our ability to briefly maintain and also operate on a limited amount of information in our mind while completing some mentally demanding tasks" (Wen et al., 2015, p. 1). Among the various conceptualisations of working memory, Baddeley's (2000) model has been particularly influential (Juffs & Harrington, 2011). It comprises four components: (a) the central executive, which manages attentional control and coordinates complex cognitive processes, (b) the phonological loop, which stores and rehearses verbal and acoustic information; (c) the visuospatial sketchpad, which stores and manipulates visual images and spatial relations; and (d) the episodic buffer, which integrates verbal, visual, and spatial information before transferring it to long-term memory. While the central executive was traditionally conceptualised as a single cognitive process, recent research has reconceptualised it to reflect its role in specific domains of human cognition and action. Wen's (2016) integrated framework of working memory in SLA, for example, categorises the central executive into three subprocesses: updating (continuously monitoring, revising, and updating incoming information), switching (flexibly alternating between tasks), and inhibition (intentionally suppressing responses when necessary).

So far, L2 experimental studies have primarily focused on the role of working memory in intentional vocabulary learning (Wen & Li, 2019), while its role in incidental learning remains relatively unexplored, particularly within the TBLT framework. A few studies have looked into the effect of working memory on incidental vocabulary acquisition from multimodal input, with Malone (2018) being one of the first to address the issue. Intermediate-level L2 learners were randomly assigned to four treatment groups, each completing reading tasks with target words embedded either two or four times, with or without aural enhancement. Working memory was measured with a nonword span task, an Automated Operation Span (OSPAN) task, and a Shape Builder task. The results revealed a significant correlation between composite working memory scores and form recognition in the group that engaged in reading-while-listening. The TwiLex Group (2024) did a replication of Malone (2018) that further explored the role of participants' L1 background. The results also showed a significant correlation between composite working memory scores (the sum of a nonword span and an OSPAN span task) and form recognition.

While Malone (2018) and the Twilex Group (2024) treated working memory as a unified construct, Montero Perez (2020) examined phonological short-term memory and complex working memory in incidental vocabulary acquisition through video viewing with upper-intermediate L2 learners. Results indicated that complex working memory, measured by a backward digit span and an OSPAN task, positively correlated with the form and meaning recognition, but phonological short-term memory, assessed through a forward digit span task, did not predict post-test scores. The study attributed the non-significant finding to participants' higher proficiency level, as previous research found that the predictive effect of phonological short-term memory diminishes with increasing proficiency (Masoura & Gathercole, 2005).

A series of studies by Teng (2023a, 2023b, 2024a, 2024b) further investigated the relationship between working memory and vocabulary gains from captioned videos. In line with Montero Perez's (2020) finding, Teng (2023a) reported that university-level EFL learners' complex working memory (measured by a reading span task), rather than phonological short-term memory (assessed using a nonword span task), significantly predicted delayed meaning recognition and recall. In addition, Teng (2023b) found that L2 young learners' phonological short-term memory, captured through a nonword span task, was a significant predictor of immediate and delayed form recognition and meaning recall, while complex working memory, assessed via an OSPAN task, did not predict learning gains. This provides additional support for the findings that phonological short-term memory plays a more facilitative role among low-proficiency learners.

Teng (2024a, 2024b) focused solely on complex working memory, operationalised through a reading span task. Teng (2024b) reported a significant effect of complex working memory on vocabulary gains across reading, listening, reading-while-listening, and viewing. Teng (2024a), however, found no significant effects of complex working memory or aptitude in the comedy and education genres, although both moderated the impact of L2 proficiency on form and meaning recognition. The researcher attributed the

non-significant finding to participants' proficiency, which might have diminished the influence of complex working memory.

Taken together, although previous research has explored the involvement of phonological short-term memory and complex working memory in L2 vocabulary acquisition from viewing, little is known about the roles of visuospatial short-term memory and other central executive functions (Li, 2024). Given that video-lecture-based tasks present information through both auditory and visual channels, further investigation is needed on how visuospatial short-term memory supports the processing of written words. Additionally, video-lecture-based tasks involve multiple operations, suggesting the need to examine how learners' ability to shift attention among task demands facilitates vocabulary processing and learning. To address these gaps, the current study employed a range of working memory tasks to look into the roles of phonological and visuospatial short-term memory, as well as the updating and task-switching functions of the central executive. Similar to research on aptitude, pursuing this line of enquiry is key, as the findings may help identify instructional conditions that support vocabulary development for learners with varying working memory capacities.

Research Questions

Against this background, the following research questions (RQs) were formed to guide the current study:

To what extent do aptitude and working memory affect

1. Learners' learning of technical words through a video-lecture-based task, as measured by offline vocabulary tests (immediate and delayed form recognition, meaning recall, and meaning recognition)?
2. Learners' visual attention to the target words during a video-lecture-based task, as reflected in their eye movements (total fixation duration, mean fixation duration, and fixation count)?

For RQ1, drawing on the overall positive effects of aptitude and working memory on incidental vocabulary learning (e.g., Montero Perez, 2020; Muñoz et al., 2024; Nagata et al., 1999), we hypothesised that these two cognitive factors would be positively linked to vocabulary gains from the video-lecture-based task. Our hypothesis for RQ 2 was nondirectional, given the limited empirical research evidence.

3. Methods

3.1. Design

The dataset for the present study is a subset of a larger project exploring how video-lecture-based tasks affect multimodal L2 processing and learning of technical vocabulary (Shi et al., 2024). The current report only focuses on parts of the design pertinent to the specific research questions being addressed. A total of 30 Chinese L2 users of English studying at a UK university participated in the study. They were enrolled in social science programmes, with most majoring in applied linguistics, teaching English to speakers of other languages (TESOL), and education. One participant was excluded because they did not complete all aptitude and working memory tests. As illustrated in Fig. 2, the remaining 29 participants completed a listening proficiency test and a vocabulary size test

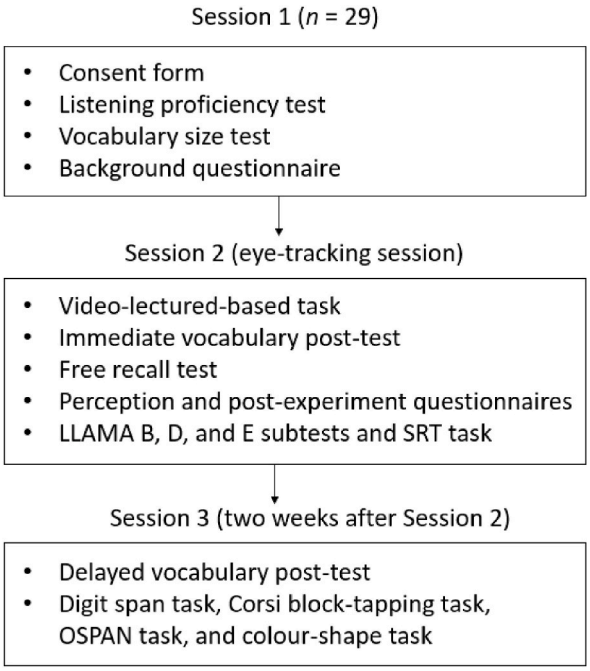

Fig. 2. Visual diagram of the research design.

(VST, Nation & Beglar, 2007) before the treatment. In the second eye-tracking session, they performed a video-lecture-based task (watching a video lecture on neurobiology while taking notes) and then took a surprise vocabulary post-test, a free recall test, a perception, and a post-experiment questionnaire, as well as a set of aptitude tests. Two weeks after the treatment, the participants completed a delayed vocabulary post-test and a battery of working memory tests.

3.2. Participants

The main study sample consisted of 27 female and 2 male participants, with ages between 21 and 28 ($M = 23.62$, $SD = 1.97$, 95 % CI [22.87, 24.37]). The large majority of participants' listening proficiency levels fell into the B2-C1 bands according to the Common European Framework of Reference (CEFR), as measured by the listening section of a Cambridge Certificate in Advanced English (CAE) test. See Appendix A for the distribution of participants in each CEFR band. The VST (Nation & Beglar, 2007) was used to obtain a vocabulary size score for each participant. The participants' mean receptive vocabulary size was 8820.69 (SD = 1345.46; 95 % CI [8308.90, 9332.48]).

3.3. Video-lecture-based task

The task asked participants to watch a Coursera lecture on neurobiology while taking notes for a friend who could not attend the lecture (see Appendix B for the task instructions). Participants were instructed to take notes by hand on the provided paper, a method they were all familiar with, in order to control for the potential effects of note-taking modalities (i.e., writing versus typing, Mueller & Oppenheimer, 2014). The lecture consisted of three short video prompts, featuring an English L1 lecturer introducing fundamental concepts in neurobiology with labelled diagrams (see Fig. 3). Each of the three videos was approximately 6 min and contained an average of 770 words, embedding four, four, and three target words, respectively. Lexical profile analysis showed that knowledge of the first 3000 most frequent word families provided a lexical coverage of 93 %, 88 %, and 94 % for the three videos, respectively. To retain task authenticity, except for trimming and combining videos to an appropriate length, no further attempt was made to modify or simplify the videos.

3.4. Target words

The selection of target words for the study followed a structured process, that is, identifying words representing main ideas in the videos using contextual information, a technical dictionary, and expert consultation (Chung & Nation, 2004). For the first video, the words *soma*, *dendrite*, *axon*, and *synapse* were chosen to represent the four basic parts of neurons. The second video introduced the brain membranes, specifically known as *meninges*, which consisted of *dura*, *pia*, and *arachnoid*. The third video explained how photoreceptor cells function under different lighting conditions, namely, *scotopic*, *mesopic*, and *photopic*. These words were confirmed to be accompanied by at least one aural definition and presented in written form as diagram labels with hand-drawn illustrations. The relevance and technical nature of the selected words were further validated through a neurobiology glossary (Purves et al., 2018) and consultation with a doctoral researcher in the field.

To ensure that participants did not have prior knowledge of the target words in either their L1 or L2, these words were piloted with 34 students from similar backgrounds and from the same university as those who would participate in the actual study using the same set of vocabulary tests (form recognition, meaning recall, and meaning recognition) as in the main study. The results confirmed that the final bank of 11 target words was known by fewer than 3 % of the pilot participants across the three tests. To control for prior topic familiarity, the participants in the main study were asked in a sign-up form whether they had previously been enrolled in any course related to psychology, neuroscience, or neurobiology. Given that the focus of this study was incidental vocabulary learning, participants were only informed that we aimed to explore multimodal lecture comprehension, both in the sign-up form and information sheet, to prevent them from consciously focusing on the target words during task performance. A debriefing session was held after the experiment to clarify the actual aims of the study to the participants. It should be noted that the methodological definition of incidental learning adopted in the current study did not prevent the participants from consciously attempting to learn the words. The academic

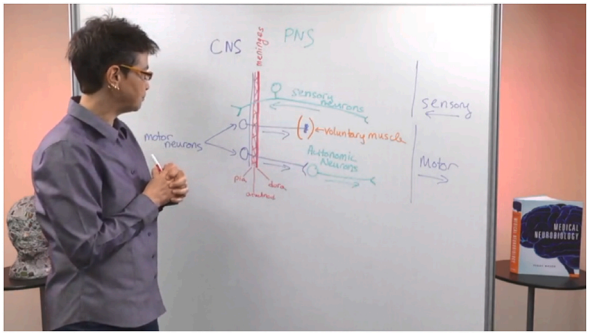

Fig. 3. A screenshot of the video prompt.

context of the video-lecture-based task might have also contributed to such conscious engagement.

Among the 11 target words, eight were nouns and three were adjectives. The average word length was 6.3 letters, ranging from three to eight letters. Each target word appeared between one to six times in the lecture and fell into the category of low-frequency words, ranked above the 11,000-word British National Corpus level. To maintain the lecture's authenticity, target word frequency of occurrence was not controlled for but was included as a covariate in the statistical analyses, given that previous research findings suggested a positive effect of frequency on incidental vocabulary acquisition (e.g., Webb, 2020). See Appendix C for a summary of target word characteristics.

3.5. Free recall test

We used the free recall test to measure participants' lecture comprehension because it closely mirrored real-life situations where learners listen to a lecture and retell the content, thus ensuring that the test had an authentic communicative purpose. Participants were asked to leave a voice message for a friend who could not attend the lecture, providing as much detail as possible about the lecture based on the notes they had taken. Note-viewing was permitted, as it was a natural part of academic listening/viewing. Participants could use their L1, L2, or a mixture of both. No time limit was set.

3.6. Vocabulary post-tests

To obtain a multi-faceted picture of incidental vocabulary development, we employed multiple measures of participants' knowledge of the target words, administered in the following order: a form recognition, a meaning recall, and a meaning recognition test. Considering that the vocabulary test had been piloted by a similar group and the results indicated that participants in the main study were unlikely to know the target words, we chose not to include a pre-test to avoid testing effects. The same test was used as immediate and delayed post-tests. The form recognition and meaning recognition tests were constructed using the software E-prime 2.0, while the meaning recall test was administered in paper-pencil format.

3.6.1. Form recognition test

Right after the participants had completed the video-lecture-based task, they took the immediate post-test. In the first, form recognition test, 11 target words and 11 distractors were randomised and presented to the participants in both written and spoken forms, one at a time. The participants then indicated whether they remembered seeing and/or hearing the words in the lecture. The distractors were neurobiology or medical terms that contained 2–3 syllables and were of the same word class as the target words. They were found to be familiar to less than 1 % of the pilot participants.

3.6.2. Meaning recall test

Following the form recognition test, participants' knowledge of the meaning of the target words was measured with a meaning recall test. The 11 target words were presented in a random order one-by-one on a slide in both written and spoken forms. Participants were asked to write down everything they knew about the meaning of the words in either their L1 or L2.

3.6.3. Meaning recognition test

A meaning recognition test was administered last to capture knowledge below the level of meaning recall. Multiple-choice items were developed for the target words, with five options: the correct definition, a definition of a target word from the same video, a definition of a target word from a different video, a definition of a distractor, and an option for "I don't know". The 11 distractors were neurobiology terms semantically related to the content of the lecture. As in the form recognition test, the target words were randomised and presented in both visual and auditory forms, and participants were asked to select the meaning closest to that of the target word. For the complete vocabulary post-test, see Appendix D.

3.7. Questionnaires

Three questionnaires were administered to the participants in paper-pencil format. We used the background questionnaire to elicit participants' demographic information. The perception questionnaire was administered to the participants immediately after they completed the free recall test. It primarily aimed to assess participants' topic familiarity. Only participants indicating a moderate unfamiliarity with the lecture topic would be included in the analyses. A post-experiment questionnaire was also administered to measure participants' knowledge of the target words prior to the experiment. Those who reported being familiar with the target words were excluded from the analyses. The questionnaires confirmed that the participants in our final sample had little knowledge of the lecture content and the target words. See Appendix E for the questionnaires.

3.8. Aptitude measures

We used three subcomponents of the LLAMA test (Meara, 2005) to measure participants' language learning aptitude that was especially relevant to incidental learning through video-based lectures, including abilities of rote vocabulary learning (LLAMA B), sound recognition (LLAMA D), and sound-symbol association (LLAMA E). This choice was based on the fact that the video-lecture-based task presented target word forms and meanings in multiple modalities and involved learners mapping new L2

forms onto new concepts. LLAMA F, which measures inductive language learning ability, was therefore excluded. The test instructions were given in the participants' L1 following the LLAMA manual. For the LLAMA D subtest, however, to prevent the participants from consciously paying attention to the linguistic forms, we adopted an incidental instruction format. That is, we only informed participants that their next task was to listen to a recording, without mentioning the subsequent sound recognition test. The recording was then played when participants indicated they were ready. The SRT task (Nissen & Bullemer, 1987) involved repeated exposure to sequences of stimuli, similar to how target word information was repeatedly presented in the lecture. Therefore, we administered the task to measure participants' implicit learning ability using the Inquisit 6 Lab (Millisecond, 2019).

3.8.1. LLAMA test

LLAMA B measured learners' ability to learn written forms of new vocabulary items by associating words with images. In the preparation phase, they were given 2 min to click on 20 images to display the names of different objects and match the objects with their names. In the testing phase, the participants were asked to associate the names of randomly presented objects with the correct image.

LLAMA D assessed learners' ability to recognise repeated patterns in spoken language. First, the participants were informed that they would listen to a short audio recording. Afterwards, they completed a surprise recognition test asking them to listen to 30 sound strings and distinguish sounds they had heard during the first listening from novel ones.

LLAMA E gauged learners' phonetic coding ability. In the preparation phase, the participants were given 2 min to click on 24 unfamiliar alphabetical symbols to listen to their syllables. They were expected to identify and remember the relationship between the sounds and the symbols (one syllable per symbol). Note-taking was allowed in this phase. Subsequently, the participants were asked to listen to 20 combinations of two-sound strings and choose their correct written representations.

3.8.2. Serial reaction time task

The SRT task presented participants with four grey boxes on a computer screen and were instructed to press a corresponding response key as fast as possible when one of the boxes turned red. They were not informed of the existence of a repeating pattern. The test started with a practice block of 36 randomly ordered trials. After the practice block, participants completed 12 training blocks, each containing 10 repetitions of a 12-item sequence. After the twelfth block, a new sequence was presented in Block 13, followed again by the target sequence in the final block. The participants' implicit sequence-learning ability was quantified as the reaction time difference between Block 13 (new sequence) and Block 12 (target sequence).

3.9. Working memory measures

We used four working memory measures to assess different constructs associated with working memory. Phonological short-term memory was gauged with Zhao's (2013) digit span task. Visuospatial short-term memory was measured using the forward Corsi block-tapping task. Executive functions of updating and task-switching were assessed with an OSPAN task (Turner & Engle, 1989) and a colour-shape task (Miyake et al., 2004), respectively. The digit span task was presented aurally, and the participants' oral repetition of presented sequences was recorded by a voice recorder. The other working memory measures were administered using the Inquisit 6 Lab (Millisecond, 2019). All instructions were given in participants' L1.

3.9.1. Digit span task

Zhao's (2013) Chinese digit span task asked participants to repeat sequences of two to nine randomly generated double-digit numbers in the same order. Each sequence had three trials, and the test stopped when participants were unable to correctly recall all sequences of the same length. Participants' digit span was determined by the longest sequence that was recalled successfully.

3.9.2. Corsi block-tapping task

The forward Corsi block-tapping task (Kessels et al., 2000) presented participants with two to nine identical blocks highlighted in different orders on a computer screen. For each trial, the participants had to click the blocks in the same order as they had been highlighted. The number of highlighted blocks increased from two to nine. There were two trials for each block length, and the test stopped if participants failed to recall two sequences of equal length. Participants' total score was calculated by counting the number of correctly repeated sequences.

3.9.3. Automated Operation Span task

The OSPAN task (Turner & Engle, 1989) required the participants to solve mathematic operations while remembering a sequence of unrelated English letters. Participants first determined whether a given answer to a mathematic equation was correct or incorrect and then remembered a letter. This was repeated until they were asked to recall all the letters in the presented order. The number of letters that the participants had to recall was defined as the set size, ranging from three to seven. The task entailed three sets for each set size. The total number of letters presented to participants in their correct order was 75, which was also the maximum score participants could achieve.

3.9.4. Colour-shape task

The colour-shape task (Miyake et al., 2004) presented the participants with coloured shapes and asked them to sort the stimuli by colour (red vs. green) or by shape (circle vs. triangle) by pressing corresponding keys as quickly as possible. Before each stimulus

appeared on the screen, the participants were given a cue letter ("C" or "S"). When the cue letter "C" was displayed, the participants had to identify whether the colour of the stimulus was red or green. When the letter "S" appeared, they needed to indicate whether the shape was a circle or a triangle. The task-switching cost was used to determine the participants' task-switching ability by calculating differences in mean reaction time between repeated and shifting blocks.

3.10. Data collection procedures

Data collection occurred through three experimental sessions. In the first session, participants signed an informed consent form and completed a background questionnaire, the listening section of the CAE test, and the VST. This section lasted approximately an hour. The second session took place individually in an eye-tracking laboratory. The participants did the video-lecture-based task once, during which their visual attention to the target words was captured by an EyeLink 1000 Plus eye-tracker sampling their right eye at 1000 Hz. Participants' notes were collected after lecture viewing, so they could not refer to their notes during the immediate vocabulary post-test. Next, they completed the immediate post-test, free recall test, perception and post-experiment questionnaires, the LLAMA B, D, and E subtests, and the SRT task, which together took approximately 2 h. Two weeks later, in the third session, a surprise delayed post-test was administered, followed by the digit span, Corsi block-tapping, OSPAN, and colour-shape tasks. This session lasted about 1.5 h.

3.11. Data coding and analyses

3.11.1. Vocabulary post-tests

We scored the form and meaning recognition tests on a binary scale (1 for correct, 0 for incorrect answers), with the maximum possible scores of 22 points for the form recognition (including 11 distractors) and 11 for the meaning recognition test. Following Gablasova (2014), we calculated the proportion of core meaning components that participants successfully recalled. For each target word, we identified three to four core components (see Appendix F), with a maximum meaning recall score of 11 points. For instance, if a participant recalled two out of three essential components, the score for that target word would be 2/3. Twenty percent of the data were coded by a second researcher, resulting in high inter-coder reliability for both the immediate (Cohen's kappa = .86) and delayed post-test (Cohen's kappa = .85).

3.11.2. Free recall

We began by segmenting the video transcripts into 338 idea units following Carrell's (1985) operational definition of an idea unit, after removing disfluency features and irrelevant content. This set the maximum possible score. The recalls were then transcribed, checked, and coded for idea units, with disfluencies and irrelevant content removed. Twenty percent of the data were coded by a second researcher, yielding high inter-coder reliability (Cohen's kappa = .90).

3.11.3. Eye-movement data

We conducted eye-movement data analyses using the EyeLink Data Viewer software. Data cleaning was first performed following recommendations in previous research (Godfroid, 2020). Dynamic Areas of Interest (AOIs) were then created for each occurrence of the targets, with adjustments made to their position and size to capture the movement of the target words on the screen. These AOIs were only activated during the time the target words were presented. The AOI size and duration for each target word are listed in Appendix C. We extracted three eye-movement measures from each AOI: total fixation duration, mean fixation duration, and fixation count. Total fixation duration was selected because it is a late measure capturing both the initial retrieval and subsequent integration of a word and has shown the strongest associations with learning in previous research (Godfroid et al., 2018). Total fixation count and mean fixation duration were included as complementary measures to provide alternative perspectives on considering the overall attention paid to the AOIs during the viewing process. Pearson intercorrelations indicated a strong and positive relationship between total fixation duration and fixation count (see Appendix G).

3.11.4. Notes

Participants' notes were analysed to reveal the type and frequencies of target word-related information recorded. Loosely drawing on Nation's (2013) framework of different aspects of word knowledge, including form, meaning, and use, we classified information about target words into three main types: L2 form, L2 meaning, and nonverbal illustration. Illustrations were included as they represent an integral part of the video-based lecture and contribute to learners' comprehension of the target words. The presence of a target word in the notes was coded as 1 (present) and 0 (absent). Similarly, if a nonverbal illustration of the target word was included, it was coded as 1; otherwise, 0. The extent to which learners recorded the meanings of target words was assessed using the scoring criteria from the meaning recall test, calculated as the proportion of core components identified for each target word, with values ranging from 0 to 1. Notably, no participants included L1 translations of the target words' forms or meanings, which might suggest unfamiliarity with the target words. Twenty percent of the data were coded by a second researcher, yielding high inter-coder reliability (Cohen's kappa = .92). See Appendix H for an example of participants' notes.

3.12. Statistical analyses

The statistical analyses were conducted with the R software (R version 4.1.1, R Core Team, 2021). To address RQ1, we constructed

a series of logistic mixed-effects models using the *lme4* package, with aptitude (the LLAMA B, D, and E subtests and SRT task) and working memory tests (the digit span, Corsi block-tapping, OSPAN, and colour-shape tasks) as fixed effects, respectively. The dependent variables were the vocabulary post-tests (immediate and delayed form recognition, meaning recall, and meaning recognition). To answer RQ2, we construct linear mixed-effects models for continuous dependent variables (total and mean fixation duration) and Poisson mixed-effects models for count data (fixation count). The fixed effects were the aptitude and working memory tests. All continuous variables were rescaled and centred to meet statistical assumptions.

Fixed effects were entered into the models simultaneously using the forced entry method, which ensured that all variables were considered based on theoretical relevance (Studenmund & Cassidy, 1987). Participant and item served as random effects in all models. Frequency of occurrence was included as a covariate and was only kept in the models when it significantly contributed to the model fit. Effect sizes for fixed effects (marginal R^2) and for the combined fixed and random effects (conditional R^2) were interpreted following Plonsky and Ghanbar's (2018) guidelines, with R^2 values $\leq .20$ and values $\geq .50$ considered as indicative of small and large effect size, respectively. Additionally, we reported odds ratio (*OR*) as an alternative effect size measure for logistic mixed-effects models, with odds ratio greater than 3 or less than .33 considered strong (Haddock et al., 1998). An alpha level of $p < .05$ was set for all tests. See Appendix I for model diagnostics.

4. Results

4.1. Preliminary analyses

The participants recalled an average of 57.20 units (16.92 %), $SD = 25.77$, 95 % CI [47.58, 66.82]. The descriptive statistics for the participants' vocabulary post-test scores, eye-movement measures, and aptitude and working memory test scores are presented in Tables 1–3. Descriptive statistics for target word information in notes are provided in Appendix J. The results indicated that participants successfully acquired target word knowledge at the levels of form recognition, meaning recognition, and meaning recall, and that they allocated substantial visual attention to the target words. Most participants recorded all the target word forms and a large proportion of their meanings. Approximately half of the participants included illustrations of the target words. We also did Pearson correlations among the aptitude and working memory tests. As shown in Table 4, a moderate positive correlation was found between LLAMA B and LLAMA D subtest scores, suggesting that, overall, the measures assessed different aspects of aptitude and working memory.

4.2. Research question 1: individual differences and vocabulary learning

The logistic mixed-effects models examining how aptitude and working memory affected vocabulary gains revealed that none of the aptitude measures was a significant predictor of post-test scores. In terms of working memory, as shown in Table 5, a significant negative correlation was found between digit span and immediate meaning recognition. Additionally, a positive relationship was observed between task-switching costs and immediate meaning recognition, indicating that lower task-switching ability was associated with higher immediate meaning recognition. A negative relationship was also identified between Corsi block span and immediate meaning recall. See Appendix K for full model results.

4.3. Research question 2: individual differences and attention allocation

The linear and Poisson mixed-effects models found no significant relationships between aptitude or working memory measures and learners' visual attention to the target words. Full model results are presented in Appendix L.

5. Discussion

5.1. Aptitude, working memory, and incidental vocabulary learning

The first research question aimed to examine the effects of aptitude and working memory on the incidental learning of technical words through the video-lecture-based task. Contrary to our hypothesis, we did not find any significant relationships between both

Table 1
Descriptive statistics for the immediate and delayed vocabulary post-test scores.

Test	*M*	*SD*	95 % CI
Immediate form recognition	16.03	2.24	[15.18, 16.89]
Immediate meaning recall	1.78	1.58	[1.18, 2.38]
Immediate meaning recognition	5.93	2.60	[4.94, 6.92]
Delayed form recognition	16.66	2.50	[15.71, 17.60]
Delayed meaning recall	.89	1.54	[.31, 1.48]
Delayed meaning recognition	5.83	2.58	[4.85, 6.81]

Note: Maximum score for form recognition = 22; meaning recall = 11; meaning recognition = 11.

Table 2
Descriptive statistics for eye-movement measures on the target words.

Eye-movement measures	*M*	*SD*	95 % CI
Total fixation duration	52668	15110	[46920, 58415]
Mean fixation duration	318.87	52.28	[298.99, 338.76]
Fixation count	166.38	44.93	[149.29, 183.47]

Note: Fixation duration measures are in milliseconds.

Table 3
Descriptive statistics for aptitude and working memory tests.

Test	*M*	*SD*	95 % CI
LLAMA B	46.38	17.52	[39.72, 53.04]
LLAMA D	34.14	13.50	[29.00, 39.27]
LLAMA E	80.00	15.58	[74.07, 85.93]
SRT	97.00	94.30	[61.13, 132.87]
Digit span	4.41	.82	[4.10, 4.73]
Corsi block span	6.28	.96	[5.91, 6.64]
OSPAN	55.28	13.20	[50.26, 60.30]
Task switching costs	213.74	253.42	[117.35, 310.14]

Note: The maximum scores are as follows: LLAMA B, D, and E subtests = 100; digit span task = 9; Corsi block-tapping task = 9; OSPAN task = 75; reaction times in the SRT and colour-shape task are in milliseconds.

Table 4
A correlation matrix of aptitude and working memory test scores.

Test	1	2	3	4	5	6	7	8
1 LLAMA B	–	.50**	−.05	−.16	−.18	.17	.10	.18
2 LLAMA D		–	.03	−.05	.16	.24	.20	−.04
3 LLAMA E			–	−.04	.05	−.22	−.22	−.36
4 SRT				–	−.10	.11	−.07	−.02
5 Digit span					–	.17	.17	−.32
6 Corsi block span						–	.21	−.24
7 OSPAN							–	−.23
8 Task switching costs								–

Note: $^{**}p < .01$.

Table 5
Results of logistic mixed-effects models.

Variable	*b*	95 % CI	*SE*	OR^a	95 % CI	*z*	*p*	R^2m	R^2c
Immediate meaning recognition – Digit span	−.68	[-1.14, −.22]	.23	.51	[.32, .80]	−2.90	.004	.07	.16
Immediate meaning recognition – Task-switching costs	1.46	[.34, 2.58,]	.57	4.30	[1.40, 13.21]	2.55	.01	.07	.16
Immediate meaning recall – Corsi block span	−.45	[-.86, −.02]	.21	.64	[.42, .97]	−2.08	.04	.06	.27

Note: [a]Odds ratio.

explicit and implicit aptitude measures and vocabulary gains at any level. This aligns with findings from Teng (2024a) but contrasts with previous studies that found a positive link between aptitude and learning gains (Muñoz et al., 2024; Nagata et al., 1999; Teng, 2022). One possible explanation for the discrepancy might be due to participants' L2 proficiency level. While participants in those studies were high school and university EFL students, participants in the current study were primarily L2 learners/users at an upper-intermediate to advanced level. As L2 proficiency increases, the influence of aptitude might diminish (Winke, 2013). Moreover, the relatively low mean scores on the immediate and delayed meaning recall tests ($M = 1.78$; $M = .89$) suggest potential floor effects, which might partly account for the lack of significant associations. This observation also echoes previous findings that L2 learners' knowledge of technical vocabulary acquired through academic input tends to be general and superficial (e.g., Gablasova, 2014).

Turning to implicit aptitude more specifically, neither the LLAMA D subtest nor the SRT task yielded significant effects. One possible explanation is that learners might have engaged more in explicit than implicit learning during the task. Specifically, given that they were asked to take notes for a friend and were aware of the upcoming free recall test, they were likely encouraged to consciously pay attention to the words or use deliberate memorisation strategies. As a result, the influence of implicit aptitude might have been limited in this context. Additionally, the SRT task did not involve form-meaning mapping (Li, 2022), potentially making it less predictive of lexical forms compared to morphosyntax structures.

In terms of working memory, we found no significant relationships between any measures and form recognition, possibly because

learners used their notes as an external storage tool for word forms. In addition, better performance on the digit span task ($OR = .51$) and the colour-shape task ($OR = 4.30$) was associated with poorer immediate meaning recognition. Similarly, a negative relationship was found between Corsi block span and immediate mearning recall ($OR = .64$). These seem to contradict our expectations and previous research findings that working memory tends to have an overall positive correlation with incidental vocabulary gains. This might also be due to the types of processing elicited by the viewing tasks. In previous studies, participants might have processed the target words less deliberately, as their viewing activities were situated in non-academic contexts. Participants in the current study, however, probably adopted more conscious processing strategies, which might have attenuated the influence of working memory.

The negative relationships might also be explained by the cognitive demands of the video-lecture-based task, which involve phonological and visual processing, as well as task-switching ability. First, we revealed a negative relationship between digit span and immediate meaning recognition, with a small effect size. Learners with stronger phonological short-term memory might efficiently record target word information, leaving extra cognitive resources available to note down illustrations as well. Consequently, they might inadvertently prioritise processing visual aids over auditory target word information, potentially leading to reduced immediate meaning recognition, given that the visual aids offered less semantic information than the auditory input.

Furthermore, we identified a strong negative effect of task-switching ability on immediate meaning recognition, as indicated by a large odds ratio (>3). Better task-switching ability might help learners allocate their attention across multiple task demands (Révész, 2012), including selecting, organising, and integrating both auditory and visual target word information (Mayer, 2014), while engaging in note-taking and note-reviewing. Given that simultaneously performing cognitively demanding tasks can exceed working memory capacity (Pashler, 1994), participants with lower task-switching abilities might have focused on fewer tasks rather than frequently switching between them, potentially as a strategy to manage a heavier working memory load. This might have facilitated more effective comprehension of the target word information, subsequently enhancing immediate meaning recognition.

Regarding the negative relationship between Corsi block span and immediate meaning recall (with a small effect size), a plausible explanation is that participants with stronger visual short-term memory might unconsciously focus on visual stimuli unrelated to target words. However, since these visuals provide limited semantic content, relying on them might take up cognitive resources that could otherwise be directed to processing target word information, which might therefore reduce meaning recall.

Finally, it should be noted that no significant relationship was found between updating ability, as measured by the OSPAN task, and vocabulary gains, contrary to findings in previous research (e.g., Malone, 2018; Montero Perez, 2020; Teng, 2023a, 2024b). This discrepancy might be because participants in the current study were encouraged to take notes, so they could offload key information onto paper and review it as needed. This allowed them to allocate more cognitive resources to understanding and integrating information in different modalities. In contrast, viewing without note-taking imposed greater cognitive demands, as learners must simultaneously notice word forms and infer their meanings from continuous verbal input while forming form-meaning associations. This might explain why the studies that did not permit note-taking (e.g., Montero Perez, 2020) found a positive correlation between updating ability and vocabulary acquisition.

Interestingly, beyond our main research question, we observed that the frequency of target words in the auditory input was negatively associated with delayed meaning recognition, contradicting previous findings that link frequency with greater vocabulary gains (Webb, 2020). One plausible interpretation is that repetition alone might not support durable learning of word meanings, as it did not necessarily provide more opportunities for varied semantic elaboration. For example, although the target word "axon" appeared five times in the commentaries, it was accompanied by limited contextual cues: " …, and then send it out through one *axon*. There is one *axon* … This *axon* can go far, far distances. And so, this can go a metre, easily. So this is an *axon*." Besides, since participants were engaged in note-taking, their attention might have been partially diverted from fully processing the auditory input, further reducing the benefits of repeated exposure.

5.2. *Aptitude, working memory, and attention allocation to L2 words*

The second research question sought to explore how aptitude and working memory affected learners' online processing of the target words. One possible interpretation of the non-significant results is that the task instructions (watching a lecture while taking notes for a friend) might have encouraged attention to the target word forms across participants, resulting in a ceiling effect in visual attention. This effect might have, in turn, obscured potential relationships between the cognitive factors and learning gains. Additionally, as nearly all participants recorded the target words, the externalisation of target word information might have reduced the roles of phonological and visuospatial short-term memory, as well as updating ability, in task performance. Beyond note-taking, the video-lecture-based task also presented target words as diagram labels, and such continuous visual accessibility might have further reduced the demands on working memory by diminishing the need to mentally maintain and rehearse the word forms.

Similarly, the external storage function of the notes might have reduced reliance on rote memory, as measured by the LLAMA B subtest. Given that the task presented written word forms that could potentially facilitate speech segmentation, the intrinsic cognitive load of the task might also have been reduced (Paas & Sweller, 2014), thereby mitigating the influence of sound recognition ability as measured by the LLAMA D subtest. Phonetic coding ability, assessed using the LLAMA E subtest, did not emerge as a significant predictor either, probably because the task prioritised comprehension of target word information over detailed phonetic analysis. Finally, the SRT task scores did not significantly predict visual attention, possibly due to the fact that the video-lecture-based task focused on learning individual word forms rather than morphosyntax structures or grammatical rules. Specifically, the learning process involved noticing written and spoken word forms, comprehending visual representations and oral explanations, and creating form-meaning associations, whereas the SRT task primarily targeted sequence learning rather than associative ability.

Finally, it is important to note that our study focused on late eye-tracking measures, which reflect both early and late stages of

processing (e.g., initial word recognition and subsequent integration of the word into the mental representation of the sentence or discourse). While these measures offer an estimate of processing effort under the eye-mind hypothesis (Just & Carpenter, 1980), their hybrid nature made it difficult to distinguish between specific cognitive processes.

6. Limitations

Before drawing our conclusions, it is important to acknowledge several limitations of our research. The first limitation concerns the relatively small sample size, which might lead to less robust results and limit their generalisability. We did not conduct a power analysis to determine the minimum number of participants because power analysis for mixed-effects models remains exploratory, particularly when multiple random effect groupings (e.g., participants and items) are involved (Huensch & Nagle, 2021). It would be worthwhile to expand the study with different lecture topics and learners from various language backgrounds and proficiency levels. Another limitation involves the potential influence of note-taking on participants' head movements, which might have impacted the eye-movement data quality. Additionally, the free-recall test could have provided additional learning opportunities, as participants were able to read their notes and were asked to verbally articulate the target words. This in turn could have affected their performance in the delayed post-test. Also, we used an older version of the LLAMA test (Meara, 2005); future research could benefit from employing more recent and validated tools.

7. Conclusion

The aim of this study was to explore the roles of aptitude and working memory in learners' processing and acquisition of technical vocabulary during a video-lecture-based task. We adopted a battery of aptitude and working memory tasks measuring various components of these constructs and quantified learners' visual attention to the target words using different eye-movement measures. Our findings revealed negative relationships between working memory measures and vocabulary gains, which contrasts with the general findings in L2 research. This discrepancy might primarily be attributed to the key component, note-taking, in our study, which likely influenced how learners processed target word information, compared to situations where note-taking was not permitted during viewing. Given the complex nature of video-lecture-based tasks, the role of note-taking was not straightforward. While it could help externalise target word information, potentially alleviating cognitive load and supporting deeper information processing, it also created a dual-task situation that competes for limited cognitive resources (Piolat et al., 2005). A pedagogical implication, therefore, is to provide instructions on effective note-taking strategies to maximise the benefits of note-taking (Rahmani & Sadeghi, 2011). Future research is needed to further examine the role of working memory in incidental vocabulary learning from video-lecture-based tasks.

Overall, the findings indicated that working memory was likely linked to various aspects of task performance, depending on which aspect learners prioritised (Li, 2024). Future research could also explore pedagogical interventions, such as task repetition or lecture navigational controls (e.g., pause/play, fast forward, and rewind), to alleviate the demands on working memory during multimodal L2 input processing. Task repetition, for example, has been found to ease cognitive demands as regards content and direct learners' attention to linguistic forms after the initial task performance (Skehan, 1998). Future studies could further enhance our understanding of multimodal L2 learning by exploring learners' individual learning strategies and modality preferences. This could be achieved through questionnaires and interviews that tap into learning styles (visual, verbal, kinesthetic, and tactile) using available learning style inventories (e.g., Dunn & Dunn, 1978), as well as their frequency of engagement with different types of multimedia academic input (e.g., written texts, audio materials, videos, and educational computer games).

CRediT authorship contribution statement

Danni Shi: Conceptualization, Methodology, Funding acquisition, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Andrea Révész:** Conceptualization, Methodology, Writing – review & editing. **Ana Pellicer- Sánchez:** Methodology, Writing – review & editing.

Declaration of competing interests

The authors declare none competing interests.

Acknowledgements

This study was funded by the *Language Learning* Dissertation Grant Program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2025.103749.

References

Baddeley, A. D. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language and learning, 70*(1), 11–47. https://doi.org/10.1111/lang.12368

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *Tesol Quarterly, 19*, 727–752. https://doi.org/10.2307/3586673

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in Language Learning aptitude* (pp. 83–118). Newbury House.

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language aptitude test*. Psychological Corporation.

Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture note-taking. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 75–92). Cambridge University Press.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System, 32*(2), 251–263. https://doi.org/10.1016/j.system.2003.11.008

Conklin, K., Pellicer-Sánchez, A., & Carroll, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dahlen, K., & Caldwell–Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal, 97*(4), 902–916. https://doi.org/10.1111/j.1540-4781.2013.12045.x

Dang, T. N. Y. (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *Tesol Quarterly, 54*(2), 436–459. https://doi.org/10.1002/tesq.552

Di Vesta, F. J., & Gray, G. S. (1973). Listening and note taking: II. Immediate and delayed recall as functions of variations in thematic continuity, note taking, and length of listening-review intervals. *Journal of Educational Psychology, 64*(3), 278–287. https://doi.org/10.1037/h0034589

Dunn, R., & Dunn, K. (1978). *Teaching students through their individual learning styles*. Reston.

Ellis, R. (1999). *Learning a second language through interaction*. John Benjamins.

Ellis, R. (2003). *Task-based language teaching and learning*. Oxford University Press.

Field, J. (2013). The cognitive validity of the lecture-based question in the IELTS Listening paper. In L. Taylor, & C. Weir (Eds.), *IELTS collected paper 2: Research in the reading and listening assessment* (pp. 391–453). Cambridge University Press. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_rr_volume09_report1.ashx.

Frase, L. T. (1970). Boundary conditions for mathemagenic behaviors. *Review of Educational Research, 40*(3), 337–347. https://doi.org/10.3102/00346543040003337

Gablasova, D. (2014). Learning and retaining specialized vocabulary from textbook reading: Comparison of learning outcomes through L1 and L2. *The Modern Language Journal, 98*, 976–991. https://doi.org/10.1111/modl.12150

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition, 21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Granena, G. (2013a). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning, 63*(4), 665–703. https://doi.org/10.1111/lang.12018

Granena, G. (2013b). Cognitive aptitudes for L2 learning and the LLAMA language aptitude test. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate attainment* (pp. 179–204). John Benjamins.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3*(3), 339–353. https://doi.org/10.1037/1082-989X.3.3.339

Huang, L., Ouyang, J., & Jiang, J. (2022). The relationship of word processing with L2 reading comprehension and working memory: Insights from eye-tracking. *Learning and Individual Differences, 95*, Article 102143. https://doi.org/10.1016/j.lindif.2022.102143

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning, 71*(3), 626–668. https://doi.org/10.1111/lang.12451

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Blackwell.

Iizuka, T., & DeKeyser, R. (2024). Scrutinizing LLAMA D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition, 46*(1), 28–50. https://doi.org/10.1017/S0272263122000559

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied Neuropsychology, 7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*, 567–587. https://doi.org/10.3138/cmlr.59.4.567

Li, S. (2022). Explicit and implicit language aptitudes. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of SLA and individual differences* (pp. 37–53). Routledge.

Li, S. (2024). Individual differences and task-based language teaching: Theory, research, and practice. In S. Li (Ed.), *Individual differences and task-based language teaching*. Routledge.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*(3), 530–566. https://doi.org/10.1111/lang.12011

Liu, D., & Lei, L. (2020). Technical vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 111–124). Routledge.

Long, M. H., & Doughty, C. J. (2009). *Handbook of language teaching*. Blackwell.

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition, 40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Masoura, E. V., & Gathercole, S. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3–4), 422–429. https://doi.org/10.1080/09658210344000323

Mayer, R. (2014). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press.

Meara, P. M. (2005). *LLAMA language aptitude tests*. Lognostics.

Millisecond. (2019). Inquisit 6 Lab [Computer software]. Retrieved from http://www.millisecond.com/.

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*(2–3), 123–142. https://doi.org/10.1016/j.actpsy.2003.12.004

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition, 42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25*(6), 1159–1168. https://doi.org/10.1177/0956797614524581

Muñoz, C., Pattemore, A., & Avello, D. (2024). Exploring repeated captioning viewing as a way to promote vocabulary learning: Time lag between repetitions and learner factors. *Computer Assisted Language Learning, 37*(7), 1744–1770. https://doi.org/10.1080/09588221.2022.2113898

Nagata, H., Aline, D., & Ellis, R. (1999). Modified input, language aptitude, and the acquisition of word meanings. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 133–149). John Benjamins.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*, 9–13. Retrieved from https://www.lextutor.ca/tests/nation_beglar_size_2007.pdf.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220–244. https://doi.org/10.1037/0033-2909.116.2.220

Pellicer–Sánchez, A. N. A., Conklin, K., Rodgers, M. P., & Parente, F. (2021). The effect of auditory input on multimodal reading comprehension: An examination of adult readers' eye movements. *The Modern Language Journal, 105*(4), 936–956. https://doi.org/10.1111/modl.12743

Peper, R. J., & Mayer, R. E. (1978). Note taking as a generative activity. *Journal of Educational Psychology, 70*(4), 514–522. https://doi.org/10.1037/0022-0663.70.4.514

Piolat, A. (2007). Effects of note-taking and working-memory span on cognitive effort and recall performance. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition* (pp. 109–124). Brill.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*(3), 291–312. https://doi.org/10.1002/acp.1086

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal, 102*(4), 713–731. https://doi.org/10.1111/modl.12509

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Mooney, A. D., Platt, M. L., & White, L. E. (2018). *Neuroscience* (6th ed.). Oxford University Press.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing [Computer software]. Retrieved from Version 4.1.1. https://www.R-project.org/.

Rahmani, M., & Sadeghi, K. (2011). Effects of note-taking training on reading comprehension and recall. *Reading, 11*(2), 116–128. Retrieved from chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.readingmatrix.com/articles/april_2011/rahmani_sadeghi.pdf.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language and learning, 62*(1), 93–132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Shi, D., Révész, A., & Pellicer-Sánchez, A. (2024). The effects of task repetition on the processing and acquisition of technical vocabulary through video-lecture-based tasks: A mixed-methods study. *Language Learning*. https://doi.org/10.1111/lang.12679

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2002). Theorising and updating. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). John Benjamins.

Studenmund, A. H., & Cassidy, H. J. (1987). *Using econometrics: A practical guide*. Little Brown.

Suárez, M. D. M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: Do proficiency and aptitude make a difference? *Language Learning Journal, 47*(4), 497–517. https://doi.org/10.1080/09571736.2019.1617768

Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition, 43*(3), 663–676. https://doi.org/10.1017/S0272263120000704

Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System, 105*, 1–12. https://doi.org/10.1016/j.system.2022.102736

Teng, M. F. (2023a). Incidental vocabulary learning from captioned video genres: Vocabulary knowledge, comprehension, repetition, and working memory. *Computer Assisted Language Learning*, 1–40. https://doi.org/10.1080/09588221.2023.2275158

Teng, M. F. (2023b). Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory. *Computer Assisted Language Learning, 38*(1–2), 206–234. https://doi.org/10.1080/09588221.2023.2173613

Teng, M. F. (2024a). Incidental vocabulary learning from captioned video genres: Proficiency, working memory, and aptitude. *Computer Assisted Language Learning*, 1–43. https://doi.org/10.1080/09588221.2024.2421517

Teng, M. F. (2024b). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System, 124*, Article 103381. https://doi.org/10.1016/j.system.2024.103381

The TwiLex Group. (2024). First language effects on incidental vocabulary learning through bimodal input. *Studies in Second Language Acquisition, 46*, 1413–1438. https://doi.org/10.1017/S0272263124000275

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press. Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*, 13–40. https://doi.org/10.3138/cmlr.53.1.13

Wen, Z. (2016). Working memory and second language learning: Towards an integrated approach. *Multilingual Matters*.

Wen, Z., & Li, S. (2019). Working memory in L2 learning and processing. In J. Schwieter, & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 365–389). Cambridge University Press.

Wen, Z., Mota, M. B., & McNeill, A. (2015). Introduction and Overview. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 1–14). Multilingual Matters.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal, 97*(1), 109–130. https://doi.org/10.1111/j.1540-4781.2013.01428.x

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2

Zhao, Y. (2013). Working memory and corrective recasts in L2 oral production. *Asian Journal of English Language Teaching, 23*(1), 57–82. Retrieved from https://muse.jhu.edu/article/537824.