The Effects of Task Repetition on Learners' Processing of Multimodal L2 Input and Acquisition of Technical Vocabulary: A Mixed-Methods Study

A DISSERTATION

Submitted to the Faculty of

Culture, Communication and Media in fulfilment

of the requirements

for the degree of Doctor of Philosophy

by

Danni Shi

University College London

2023

Declaration Page

I, Danni Shi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

During the COVID-19 pandemic era, video-presented lectures have become an essential component of university education. Video recordings of lectures are often made available to students either following a live session or as part of asynchronous curriculum provision. Both scenarios allow for repeated viewing of lectures while providing instructional input in multiple modalities. Given the increased importance of learning through video lectures, there is a clear need to understand how second language (L2) learners process content and language during lecture viewing, and how the acquisition of linguistic forms may proceed in this multimodal context. Little research, however, has investigated the role of video-based lectures in L2 development, and even less research has looked into how repeated lecture viewing may affect L2 learners' processing of multimodal L2 input and vocabulary acquisition. Situated in the context of task-based language teaching, the current study aims to fill these research gaps.

The study adopted a mixed-methods design, including 75 Mandarin users of L2 English at a UK university as participants. They were allocated to three groups using stratified random assignment based on the results of a pre-administered listening proficiency test. The control group (n = 30) performed a lecture-viewing task once, whereas the repetition group (n = 30) did the same task three times. The task asked the participants to watch a neurobiology lecture while taking notes. The lecture featured an instructor introducing fundamental concepts of neurobiology in front of a whiteboard with labeled diagrams. Eleven key technical terms shown in the diagrams were selected as target items. The participants' visual attention to the instructor, diagrams, and target items during each viewing was captured using an eye tracker. Immediately after viewing, both groups were asked to complete an unannounced vocabulary post-test measuring their knowledge of the target items, followed by a free recall test assessing their lecture comprehension. A delayed vocabulary post-test was administered two weeks after the experiment. The stimulated recall group, on the other hand, performed the task once (n = 5), twice (n = 5), or three times (n = 5), respectively. The stimulated recall participants did not complete any vocabulary post-tests or the free recall test. Instead, they were asked to describe their thought processes during their last task performance, prompted by their notes and recordings of their eye movements during lecture viewing. The data analysis involved triangulation results from (a) eye-gaze recordings, (b) stimulated recall comments, (c) vocabulary test scores, and (d) free recall scores.

Results from mixed-effects statistical models revealed that task repetition had a significant positive effect on learners' vocabulary gains. During repeated task performance, learners' visual attention to the instructor increased, whereas their visual attention to the diagram as well as the target words declined. A negative relationship was identified between learners' attention allocated to the instructor or the diagram and their lecture comprehension. No significant difference, however, was found between the amount of attention allocated to the target words and post-test scores. Results of qualitative analyses showed that task repetition led to lower reliance on higher-level listening processes but more listening/viewing strategies. In addition, learners noticed more specific aspects of the target items during repeated viewing. The results are discussed in terms of Schmidt's (2001) Noticing Hypothesis and models of L2 listening processes, L2 listening strategies, and multimedia learning.

Impact Statement

This research was motivated by the widespread availability of video lectures that allow repeated viewing and the importance of repeated viewing in facilitating learners' processing of multimodal L2 input and acquisition of linguistic knowledge. The results of the investigation have important theoretical, methodological, and practical implications.

First, the results of the current study provide novel insights into the processes underlying video-based lecture comprehension and vocabulary acquisition from repeated exposure to multimodal L2 input, drawing on the Noticing Hypothesis (Schmidt, 2010) and models of L2 listening processes (Field, 2013), listening strategies (Vandergrift & Goh, 2012), and multimedia learning (Hegarty, 2014; Schnotz, 2014). These cognitive models appear to be useful theoretical starting points for researchers to investigate L2 multimodal comprehension and vocabulary learning from viewing. The results also contribute to our understanding of how learners allocate attention to different aspects of visual input (i.e., social cues, written verbal, and pictorial information) and how they strategically use the visual input to achieve lecture comprehension. This provides empirical evidence in support of an expansion of the construct of L2 academic listening by including the ability to understand visual information.

Second, the methodological innovation of the current research lies in triangulating quantitative (online eye-movement data and offline vocabulary tests) and qualitative data (stimulated recall comments), allowing for the painting of a richer and fuller picture of L2 learners' viewing behavior and associated cognitive processes underlying multimodal comprehension and vocabulary acquisition. This innovative approach, which combines multiple types of data, helps to offset the limitations of each research method. Another methodological novelty involves using dynamic Areas of Interest (AOIs) to capture learners' visual attention to moving objects with greater precision, thus providing a more fine-grained analysis of learners' attentional processes during viewing. In short, the study presented in this thesis demonstrates the benefits of triangulating data from multiple sources and using dynamic AOIs as a research tool, which can guide researchers' methodological decisions in the area of multimodal processing and vocabulary learning.

Finally, the results of the investigation are of important pedagogical value for classroom practitioners and L2 learners. The results of the current study provide empirical support that task repetition can help learners deal with cognitive demands imposed by having to process multimodal L2 input in real time, thus enabling learners to notice target linguistic forms and integrate information presented in different modalities. Considering that task repetition is a useful task implementation factor for L2 learners to develop their linguistic knowledge, language teachers can incorporate repeated viewing into instructional materials and lesson plans (e.g., designing activities that involve repeated exposure to linguistic forms) to consolidate and strengthen students' L2 lexical knowledge over time. The results of the current study also reveal a facilitating effect of nonverbal social cues (e.g., a speaker's eye contact, facial expressions, and gestures) on learners' acquisition of technical vocabulary, advocating that language teachers can create a supportive and engaging classroom environment using nonverbal signals, such as directional gaze, to draw students' attention to the target linguistic forms that they wish their students to master.

Acknowledgments

The journey of pursuing this Ph.D. is a mixture of joy and pain. While it is undoubtedly fulfilling in terms of intellectual stimulation and career opportunities in academia, it can also be overwhelming, stressful, and isolating. I would not have been able to complete this journey without the support that I received from many people.

First, I would like to express my deepest gratitude to my principal supervisor, Professor Andrea Révész, for her continuous guidance and encouragement throughout the journey. When I was her M.A. student, she saw potential in me and motivated me to step on the path toward a doctoral degree. During the most difficult months of the COVID-19 pandemic, she ensured that I felt supported and connected via virtual meetings. This allowed me to smoothly resume face-to-face data collection after working from home for a year and a half. Over the past five years, her mentorship has been invaluable in fostering my growth as an independent researcher.

In addition, I would like to extend my sincere appreciation to my former subsidiary supervisor, Dr. Ana Pellicer-Sánchez, whose expertise in eye-tracking has been indispensable in shaping the direction of my doctoral research. Conducting an eye-tracking study could not be as painless as she made this for me, many thanks to her constructive feedback and comments on my upgrade report, presentations, and drafts. My sincere gratitude also goes to my subsidiary supervisor, Dr. Kazuya Saito, for his wealth of experience in presenting at academic conferences and publishing journal articles, which has inspired me throughout my studies.

Apart from my supervisors, I would like to express my gratitude to other members of the Department of Culture, Communication, and Media: Dr. Talia Isaacs for sharing with me her experience as an early-stage researcher and polishing my teaching, mentoring, and research skills; and Professor John Gray for his support during the pandemic and prompt responses to process my face-to-face data collection application. I would also like to thank my Ph.D. colleagues, Yoojin Chung and Dr. Andi Wang, for their peer support. I am particularly grateful to all those who participated in my experiment. Without their contribution and generosity, my research would not have been possible.

I also wish to express my appreciation to my Ph.D. examiners, Dr. Vita Kogan and Professor. Luke Harding, for their time and effort in examining my research. Their insightful comments and suggestions have greatly contributed to the refinement of this thesis.

Last but not least, I am deeply grateful to my family, especially my mother 杨爱红. The contributions that she has made to my academic journey are immeasurable. I could not have come this far without her love and support. I would also like to extend my special thanks to my grandmother 王彩霞 and father 施江鸿.

I dedicate this thesis to my grandfather, 杨仲吉, who did not witness this journey but has always been my light in the darkness. I am forever indebted to him for his guidance.

施莟旎

July 2023

Dissemination

1. Book chapter and peer-reviewed journal

Révész, A., & Shi, D. (in prep). Learning through listening. In E. Wagner, A. Batty & Galaczi, E. (Eds.), The *Routledge handbook of second language acquisition and listening*. Routledge.

Shi, D., Révész, A., & Pellicer-Sánchez, A. (under revision). The Effects of task repetition on the processing and acquisition of technical vocabulary through video-lecture-based tasks: A mixed-methods study. *Language Learning*.

2. Conference presentation

Shi, D. (2023, August 30 – September 2). *The effects of task repetition on second language learners' phonological acquisition of technical vocabulary through multimodal academic lecture viewing*. European Second Language Acquisition (EuroSLA) Conference, Birmingham, UK.

Shi, D. (2023, March 18–21). *The role of aptitude and working memory in second language learners' processing of multimodal academic input and acquisition of technical vocabulary*. Annual Meeting of the American Association for Applied Linguistics (AAAL), Portland, USA.

Shi, D. (2022, August 24–27). *The role of language aptitude and working memory in the acquisition of technical vocabulary through multimodal academic lecture viewing*. EuroSLA Conference, Fribourg, Switzerland.

Shi, D. (2022, July 18–19). *The effects of repeated viewing of video-based lectures on advanced L2 learners' processing and acquisition of technical vocabulary*. British Association for Applied Linguistics (BAAL) Vocabulary Special Interest Group

Annual Conference, Exeter, UK.

Shi, D. (2022, March 19–22). *Learning specialized vocabulary from repeated viewing of video-based lectures*. Annual Meeting of the AAAL, Pittsburgh, USA.

Shi, D. (2021, June 11). *The effect of repeated viewing of video-based lectures on learners' processing of multimodal L2 input*. Eye-tracking in SLA Research: A Global Engagement Seminar.

Shi, D. (2021, March 20–23). *The effect of task repetition on second language listening in multimedia environments*. AAAL Virtual Conference.

Shi, D. (2019, June 20). *The effects of task repetition on second language listening and incidental vocabulary learning*. UCL Institute of Education Doctoral Summer Conference, London, UK.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOI	Areas of interest
BNC	British National Corpus
CAE	Cambridge Certificate in Advanced English
CALL	Computer-assisted language learning
CEFR	Common European Framework of Reference for Languages
CI	Confidence interval
CLIL	Content and Language Integrated Learning
CLT	Communicative language teaching
COVID-19	Coronavirus disease
EEG	Electroencephalography
EFL	English as a foreign language
ESL	English as a second language
GPC	Grapheme and phoneme correspondence
IELTS	International English Language Testing System
L1	First language
L2	Second language
LFT	Lateral flow test
OR	Odds ratio
PGR	Postgraduate research
POS	Part of speech
ROI	Region of interest
RQ	Research Question
SD	Standard deviation

SLA	Second language acquisition
TBLT	Task-based language teaching
TESOL	Teaching English to speakers of other languages
TOEFL	Test of English as a Foreign Language
VKS	Vocabulary Knowledge Scale
VST	Vocabulary Size Test
WAF	Word Associates Format

Chapter 1 Introduction

1.1 Background and Rationale of the Study

With the growing availability of multimedia and digital technologies, online education has steadily increased in popularity over the last decade (Seaman et al., 2018). Video-based lectures, an important component of online instruction (Means et al., 2009), are often freely available on e-learning platforms (e.g., Coursera, FutureLearn, LinkedIn Learning, etc.), presented by experts with the aim of rendering various academic topics accessible to a general audience. The COVID-19 pandemic has also fueled the proliferation of online learning, making video recordings of lectures an essential component of university education. Regardless of the mode of delivery (i.e., pre-recorded or synchronous), video-based lectures often capture an instructor's presence, providing a sense of social connection between the instructor and learners, which may motivate learners to maintain focused and actively engage in learning processes (Mayer, 2014a). In addition to an instructor's social cues (e.g., eye contact, facial expressions, and gestures), other types of visual aids (e.g., diagrams, written annotations, and photos) are commonly used in video-based lectures alongside aural commentaries to help illustrate technical terms and concepts.

Considering that such visual information is an integral component in most academic settings, L2 researchers have become interested in how learners interact with visuals during lecture viewing, particularly in the field of assessing L2 listening. While a few studies have investigated the amount of time test-takers spent on videos (e.g., Ockey, 2007; Wagner 2007, 2010), more recent work has employed eye-tracking methodology to explore test-takers' attention allocation to different types of visuals (i.e., visuals semantically relevant to the verbally presented information and visuals presenting the context of the spoken discourse) (Suvorov, 2015) and to speakers' nonverbal communication cues (i.e., lip movements, facial expressions, hand gestures, and body gestures) (Batty, 2020) in video-based tests. Nonetheless, scant attention has been directed toward how learners allocate attention to specific components (i.e., diagrams, written words, and the instructors' nonverbal social cues) of a video-based lecture, and what cognitive processes learners engage in to achieve multimodal comprehension. Our understanding of the relationship between attention allocation to these visual components and comprehension is also limited. These are important areas for research, as the results of the investigation could contribute to our understanding of the construct of academic listening, which has been argued to include the ability to process visual information (Ockey, 2007).

Video-based lectures also play an important role in the development of L2 learners' vocabulary knowledge, especially for those studying at university level and taught through the medium of an L2 (Vidal, 2003, 2011). Through lecture viewing, learners have the chance to repeatedly encounter technical terms of their field of study in addition to nonspecialized infrequent words in academic genres (Parry, 1991). As a result, learners may incidentally acquire new technical words representing new subject knowledge while attending to the overall meaning of an academic lecture. This process is considered more difficult than learning novel word forms alone, as it involved simultaneous learning of new concepts and associated L2 forms (Liu & Lei, 2020; Nation, 2013; Schmitt, 2010). Despite being frequently reported by students in L2-medium education as one of the greatest challenges (e.g., Evans & Morrison, 2011), little attention has been directed toward the contribution of academic lectures to the development of technical vocabulary. Although several studies have examined the acquisition of various types of vocabulary, including technical vocabulary, from lectures (e.g., Dang et al., 2021; Vidal, 2003, 2011), no study has been conducted to specifically explore the simultaneous learning of content and lexical knowledge from lecture viewing.

Given the ephemeral nature of viewing and the difficulty in acquiring technical words, it appears important to explore ways to facilitate learners' processing of multimodal input and vocabulary acquisition through L2 pedagogy. Among various pedagogical options, the potential of task repetition seems particularly worthwhile to investigate. First, situated in the context of task-based language teaching (TBLT), video-based lectures lend themselves well to designing pedagogic tasks. Tasks are defined as activities where the primary focus is on meaning, often involving a "gap" that prompts learners to rely on their own linguistic and nonlinguistic resources to communicate, and resulting in nonlinguistic outcomes (Ellis, 2009). These criteria are naturally met by lecture-viewing tasks, such as taking notes based on a video lecture. Hence, during a lecture-viewing task, L2 learners are able to engage in cognitive processes similar to those involved in most academic situations, allowing them to develop cognitive and linguistic skills that are transferable to real-life communication.

In addition, recordings of lectures are made increasingly available to students in real-life academic settings for repeated listening/viewing, which may direct their attention to additional information and facilitate the integration of input presented in different modalities that students might not be able to process fully during the first viewing. Furthermore, a large body of research has reported positive effects of task repetition on oral production (see Bygate, 2018 for a review), driven by the assumption that repeating a task allows learners to allocate increased attentional resources to linguistic forms, thereby facilitating the acquisition of L2 linguistic forms (Bygate, 1996). A few studies have also found task repetition useful for enhancing learning of lexis through listening (Ellis & Chang, 2016; Shintani, 2012a) and viewing television episodes (Majuddin & Siyanova-Chanturia, 2021). Little is known, however, about the impact of repeated lecture viewing on learners' vocabulary learning and ways in which task repetition may affect the processing of multimodal L2 input, including technical words, during viewing. The relationships between learners' processing of multimodal materials and both lecture comprehension and vocabulary acquisition also remain unexplored. These areas are of importance for research, as they provide the basis for teaching and learning technical vocabulary.

To obtain deeper insights into the complex nature of multimodal comprehension and vocabulary learning from multimodal input, L2 researchers have advocated for the combination of multiple data sources in their investigations. To be more specific, the use of eye-tracking with stimulated recall data may help gain information not only about learners' visual attention but also about their conscious thought processes during exposure to multimodal input (e.g., Jung & Révész, 2018). Additionally, these process measures would ideally be combined with comprehension tests and tests assessing different aspects of lexical mastery to obtain a more comprehensive understanding of the relationship between processing and development. These types of triangulations have already been found valuable in a few studies. Godfroid and Schmidtke (2013), for example, triangulated data from eye-movement recordings, retrospective verbal reports, and vocabulary test scores to examine how attention and awareness contributed to incidental vocabulary learning from L2 reading. Such mixed-method research is also warranted to obtain a fine-grained picture of the processes underlying multimodal comprehension and vocabulary learning from viewing.

1.2 Aim of the Thesis

To address the research gaps stated above, the current study investigated the extent to which repeating a multimodal lecture-viewing task affects: (a) learners' visual attention to the lecture instructor and labeled diagrams; (b) learners' cognitive processes underlying multimodal comprehension; (c) the relationship between learners' visual attention to lecture components and their lecture comprehension; (d) learners' incidental acquisition of technical words from lecture viewing; (e) learners' visual attention to technical words; (f) learners' awareness of technical words; and (g) the relationship between learners' visual attention to technical words and their vocabulary gains. In response to recent calls to triangulate data sources to obtain a fuller picture of attentional and acquisitional processes (e.g., Révész, 2021), a mixed-methods study was conducted. The quantitative component involved using online eye-movement data to capture learners' visual attention to a lecture instructor, diagrams, and target words during viewing, as well as administering an offline free recall test and vocabulary post-tests to measure their lecture comprehension and gains in the knowledge of the target words, respectively. Meanwhile, the qualitative component consisted of eliciting verbal reports from learners regarding their thought processes during task performance, aiming to explore the way that they engaged in multimodal comprehension and the level and source of learners' awareness of the target items. Results from the quantitative and qualitative analyses were then triangulated to provide a fuller picture of learners' cognitive processes underlying L2 viewing and incidental acquisition of technical vocabulary from viewing.

1.3 Overview of the Thesis

The remainder of the thesis includes five chapters. Chapter 2 provides a review of pertinent literature by first introducing the roles of input, attention, and awareness in second language acquisition (SLA), followed by an overview of oft-cited L2 listening models and models of multimedia learning. The chapter then proceeds with a discussion of incidental acquisition of technical vocabulary and empirical evidence on acquiring technical terms from L2 viewing. The chapter continues with the theoretical and empirical foundations for the pedagogical intervention employed in the study, namely, task repetition. It ends with an introduction of measures of learners' cognitive processes. Chapter 3 outlines the methodology adopted for the current study. It starts with a description of the mixed-methods design, research ethics, participants, materials, and instruments. The procedures related to data collection and analyses are then reported, followed by an overview of the statistical analyses. Chapter 4 presents the results of preliminary analyses to ensure the reliability of the instruments. Next, the results obtained from the quantitative and qualitative analyses are presented. Finally, Chapter 5 provides a discussion of the findings in relation to the research questions and hypotheses. After that, theoretical, methodological, and pedagogical implications of the current study are discussed. This chapter concludes by reflecting on research limitations and suggesting possible directions for future research.

Chapter 2 Literature Review

Chapter 2 of this thesis provides a review of previous theoretical and empirical work relevant to this research. First, the roles of input, attention, and awareness are addressed drawing on previous work in SLA and cognitive psychology. The next section provides a review of oft-cited cognitive models of L2 listening and listening strategies. This is followed by an overview of models of multimedia learning, leading to the description of the integrated model of multimodal lecture comprehension used in the present study. Then, the chapter turns to a discussion of incidental acquisition of technical vocabulary, followed by a review of relevant empirical studies on how learners acquire technical words from L2 reading and viewing. The chapter continues by considering previous work on task repetition, a pedagogical intervention employed in the current study. The theoretical underpinnings of task repetition are introduced before an overview of empirical studies on the effects of repetition on listening processes and vocabulary gains. The final part of the chapter is dedicated to a discussion of measures of learners' cognitive and attentional processes, including the use of verbal reports and eye-tracking technology to investigate multimodal processes. The chapter ends with a review of pertinent eye-tracking and mixed-methods studies on learners' processing of multimodal input and vocabulary acquisition.

2.1 Roles of Input, Attention, and Awareness in SLA

The prominent role of input in L2 development has long been acknowledged by SLA researchers (e.g., Ellis, 1994; Krashen, 1985, 1994; Long, 1985, 1996). Advocators of the natural approach, such as Krashen, have argued that L2 acquisition occurs incidentally when learners are able to comprehend the input to which they are exposed. According to Krashen's (1985) Input Hypothesis, being exposed to comprehensible input

– defined as input containing language structures slightly beyond a learner's current level of competence (i + 1) – is a sufficient condition for acquisition to take place. However, the hypothesis encountered much criticism for its exclusive focus on the role of input (e.g., Robinson, 1995; Schmidt, 1990; Swain, 1985; VanPatten, 1996): researchers have argued that the mere provision of input may not guarantee successful L2 learning (e.g., Doughty & Williams, 1998; Schmidt, 1990), given that learners may not pay attention to input or may not be able to process everything presented in it (Sharwood Smith, 1993). Krashen's Input Hypothesis, therefore, has been superseded by alternate theories on the roles of attention and awareness in SLA (Loewen, 2015).

2.1.1 Schmidt's Noticing Hypothesis

In response to Krashen's Input Hypothesis, the late 1980s and early 1990s witnessed a surge of theoretical discussion over the role of consciousness in learners' L2 development. Schmidt was one of the pioneers to emphasize the importance of consciousness in adult SLA. Schmidt (1990, 1994, 1995, 2001) explained that there are two levels of awareness: awareness at the level of noticing and awareness at the level of understanding. Noticing, a lower level of awareness, is isomorphic with focal attention and results in the registration of surface-level features and item learning. On the other hand, awareness at the level of understanding is associated with the establishment of rules and principles. Schmidt (1990) emphasized that awareness at the level of noticing is a necessary condition for learning to occur, while understanding plays a facilitative role in L2 learning. Later, Schmidt (1994) weakened this original version of the Noticing Hypothesis, suggesting a facilitative rather than necessary role for noticing in L2 learning, as a response to empirical findings in SLA and cognitive psychology that some learning is possible in the absence of awareness (e.g., Williams, 2009).

2.1.2 Tomlin and Villa's Model of Attention

While the term noticing is defined by Schmidt as attention plus awareness, Tomlin and Villa (1994) proposed a model that disentangled the two constructs. In their view, there are three components of attention: alertness, orientation, and detection. Alertness refers to "an overall, general readiness to deal with incoming stimuli or data" (p. 190). The second function, orientation, refers to "the specific aligning of attention ('orienting') on a stimulus" (p. 191), which may have facilitative or inhibitory effects depending on whether input occurs as expected or not. Both alertness and orientation may increase the possibility that detection occurs, but neither of these two functions is necessary. The last function, detection, is "the process that selects, or engages, a particular and specific bit of information" (p. 192). Tomlin and Villa argued that detection is necessary for further processing of the input, and language acquisition must take place at this level. Awareness, however, is not a necessary condition for L2 learning to occur, as none of the key components of attention (i.e., alertness, orientation, or detection) requires awareness.

2.1.3 Robinson's Model of Attention and Memory

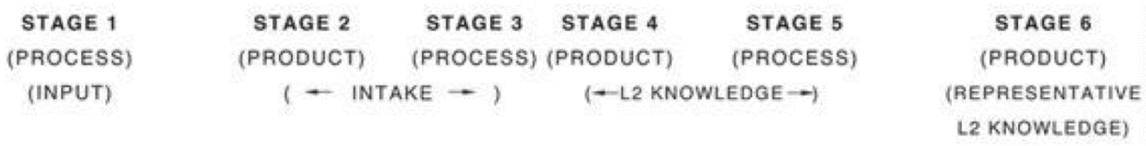
To reconcile the central difference between Schmidt's Noticing Hypothesis and Tomlin and Villa's model, Robinson (1995, p. 296) redefined noticing as "detection plus rehearsal in short-term memory, prior to encoding in long-term memory." In Robinson's (1995, 2003) model of attention and memory, learning begins with the detection of stimuli accompanied by activation of short-term memory, followed by rehearsal to store the stimuli long enough to reach the level of awareness. As a result of rehearsal, a mental trace is left in long-term memory, and input transforms into intake. On the whole, Robinson's (1995, 2003) model incorporates Tomlin and Villa's (1994) notion of detection but shares Schmidt's position that detection without awareness is insufficient to trigger noticing and subsequent L2 learning. However, Robinson and Schmidt hold different views on the cognitive mechanism underlying the process of consciousness. While Schmidt suggested that focal attention in itself implicates awareness, Robinson argued that apart from detection within focal attention, memory processes (e.g., maintenance and elaborative rehearsal) are also necessary to trigger conscious attention.

2.1.4 Leow's Model of the L2 Learning Process in Instructed SLA

More recently, synthesizing various models of attention and awareness in SLA, Leow (2015) proposed a model of L2 learning processes, in which attention is highlighted as a crucial cognitive resource at the initial stage of learning. As presented in Figure 1, there are three processing stages: an input processing, an intake processing, and a knowledge processing stage. The input processing stage is subdivided into three phases based on the level of attention involved (i.e., peripheral, selective, and focal): attended intake, detected intake, and noticed intake. According to Leow, attended intake is the product of peripheral attention, and is most likely to decay without further storage or processing in working memory. Detected intake results from selective attention to input, accompanied by a low level of processing, which is in line with Tomlin and Villa's (1994) notion of detection. When the input is cognitively registered with focal attention, combined with a low level of awareness, it is converted into noticed intake. This is equivalent to Schmidt's (1990) notion of noticing.

At the stage of intake processing, preliminary intake (i.e., attended, detected, and noticed intake) is subjected to data-driven and conceptually-driven processing. While data-driven processing involves encoding and lodging incoming intake in the L2 development system at a lower level of cognitive effort, conceptually-driven processing is accompanied by a higher level of awareness, including the conscious encoding or decoding of linguistic information. In the final stage, the knowledge processing stage, learners constantly monitor and modify their production, use available feedback to confirm or disconfirm their L2 knowledge, and utilize their own output as additional input. At this stage, the level of awareness, depth of processing, together with the ability to activate knowledge play an important role in the development of learners' L2 system.

Figure 1 Model of the L2 Learning Process in Instructed SLA (Adapted from Leow, 2015, p. 242)



2.1.5 A Taxonomy of External and Internal Attention

While L2 researchers have conceptualized the construct of attention as a unitary system comprised of multiple mechanisms (e.g., Tomlin & Villa, 1994), some cognitive psychologists view attention as multiple attentional systems (e.g., Chun et al., 2011). In an oft-cited taxonomy, Chun et al. (2011) made a distinction between external and internal attention based on the types of information that attention operates over. External attention is associated with perceptions and can be triggered by external stimuli, such as changes in modalities (i.e., vision, hearing, touch, smell, and taste), spatial locations, and time. On the other hand, internal attention selects, modulates, and maintains internally generated information, including the contents of working memory, long-term memory, task rules, and response selection. Effects of internal attention, for example, can be observed when manipulating the demands of a task that elicits learners' responses to stimuli. Although it is not the aim of the present study to compare the effects of manipulating external and internal attention on learning outcomes, the distinction between the two attentional systems may expand our understanding of the role of attention in L2 development.

2.1.6 Summary

To sum up, following Schmidt's work, researchers have developed various models to conceptually distinguish the constructs of attention and awareness. While Tomlin and Villa (1994) suggested that awareness does not play a central role in the input-to-intake stage, Schmidt (1994, 1995, 2001), Robinson (1995, 2003), and Leow (2015) regarded awareness as necessary for input to be converted to intake. These theoretical models have reached a consensus that attention to linguistic forms is a precondition for learning to take place. Building on the discussion of attention and awareness, researchers sought to understand how attentional mechanisms contribute to successful input processing and how learners' attention can be drawn to important linguistic forms in the input. Apparently, both L2 reading and listening are important input sources, potentially serving a dual purpose – to provide opportunities for L2 learners to develop their processing ability and to facilitate the acquisition of new linguistic knowledge when learners attend to and process the input. However, while L2 reading has been extensively researched, L2 listening has received little attention and remains underexplored and less understood (Vandergrift, 2007). In order to understand the role of attention and awareness in L2 listening, it is first necessary to take a closer look at how listening is defined and what processes are involved in listening comprehension.

2.2 L2 Listening Processes

Although various definitions of listening have been proposed over the years, there is no generally accepted definition of either first language (L1) listening or L2 listening (Wolvin & Coakley, 1996). Earlier definitions of L2 listening skills focused solely on what listeners do with auditory input (e.g., Lado, 1961), ignoring non-verbal elements involved in listening processes. More recent listening research appears to acknowledge the important role of visual information in L2 listening processes (e.g., Batty, 2020;

Ginther, 2002; Gruba, 1997; Ockey, 2007; Suvorov, 2015, 2018; Wagner, 2007, 2010). Lynch and Mendelsohn (2010, p. 180), for example, described listening as "making sense of spoken language, normally accompanied by other sounds and visual input, with the help of our relevant prior knowledge and the context in which we are listening." As Field (2013) pointed out, "visual information is not additional or supplementary of auditory input; it forms an intrinsic part of a listening event" (p. 115). Therefore, a more expansive definition that includes the ability to process visual information of L2 listening was adopted in the current study. To identify a model that can serve as a framework for analyzing L2 learners' listening processes, the following sections describe cognitive models of L2 listening that are most frequently cited in the literature on L2 listening teaching and testing, including Anderson's (1985) cognitive model, Rost's (2011) model of listening processes, Vandergrift and Goh's (2012) model of L2 listening comprehension, and Field's (2013) L2 listening model. The role of strategy use in listening comprehension is also discussed.

2.2.1 Anderson's (1985) Cognitive Model

Anderson's (1985) cognitive model for language comprehension is applicable to both reading and listening comprehension. Anderson identified three mental stages of processing in the comprehension of written and oral information: perceptual processing, parsing, and utilization. In the case of auditory information, perception is the lowest stage of language processing and involves recognizing sounds and segmenting those sounds into words. In the parsing stage, learners relate the perceived words to their knowledge with the help of their syntactic and semantic knowledge. Basic units obtained from this stage are propositions or chunks of information. In the utilization stage, learners combine newly-parsed propositions with existing knowledge to comprehend the entire meaning of an acoustic message. Anderson's model does not clearly explain how text modality (i.e., aural or written) affects language processing, thus the model might overlook some cognitive behaviors specific to listening.

2.2.2 Rost's (2011) Model of Listening Processes

Rost's (2011) model of listening processes offers a more comprehensive account of L2 listening processes, consisting of four types of processing: neurological, linguistic, semantic, and pragmatic processing. First, neurological processing refers to the neurological activities underlying all other types of processing. It involves physical and neurological processes associated with hearing and listening, such as converting mechanical sound signals to auditory perceptions and the processes of arousal, orientation, and focus. The next stage in Rost's model is linguistic processing. When the speech signal reaches the brain, listeners group the speech into units of spoken language (i.e., intonation units or pause units) that can be further processed within short-term memory. The following step is to recognize words and phrases, which is considered an automatic process for L1 speakers and expert L2 learners. Competent word recognition requires the activation of lexical knowledge associated with the linguistic items identified. This step is followed by syntactic parsing, involving translating the incoming speech into syntactic representations. Syntactic processing takes place at both sentence and discourse levels, aided by pragmatic and intertextual knowledge, as well as familiarity with formulaic language and semantic roles.

Importantly, Rost argued that non-verbal cues should be integrated into linguistic processing. He classified two basic types of visual signals: exophoric and kinesic signals. Exophoric signals function as references for spoken texts and are important for text interpretation; they include drawings or written texts on a whiteboard, which are crucial in academic lectures. Kinesic signals are body movements that a speaker makes while delivering the text. The most commonly occurring ones are baton signals (i.e., hand and head movements), directional gaze (i.e., eye movements used to direct listeners to exophoric references), and guide signals (i.e., systematic gestures and movements of any part of the body).

While linguistic processing is conceptualized in a bottom-up fashion, the next stage, semantic or top-down processing, covers aspects that allow listeners to link linguistic information to their prior knowledge and personal experience. It involves distinguishing new from old information, activating relevant schemata, inferring meaning on the basis of what is explicitly stated in the text, and updating memory representations guided by the previous semantic processes. The final stage of Rost's model is pragmatic processing, which encompasses the evaluation of the speaker's intention against the listeners' expectations, the activation of information about the social status of the speakers, and the integration of contextual information. Therefore, pragmatic processing enables listeners to supply interactive responses while listening and provide substantive responses in reaction to the speakers' message.

2.2.3 Vandergrift and Goh's (2012) Model of L2 Listening Comprehension

Vandergrift and Goh's (2012) model of listening comprehension highlights the role of metacognitive processes in L2 listening. Metacognition is defined as the "ability to think about own thinking or cognition, and, by extension, to think about how we process information for a range of purposes and manage the way to do it" (p. 83). Vandergrift and Goh distinguished between four components of metacognition: planning, monitoring, problem-solving, and evaluating. Planning occurs when listeners prepare to listen and establish necessary conditions for successful listening. Monitoring refers to the processes by which listeners evaluate comprehension and make corresponding adjustments. Problem-solving takes place when learners adjust their listening approach and activate strategies to solve listening problems. Finally, evaluating involves reflecting on the difficulties the listener experienced and on the results of their problem-solving efforts. Vandergrift and Goh's model is informed by Anderson's (1995) cognitive model and Levelt's (1989) speech-production model, consisting of listening and speaking processes to explain both one-way and interactive listening. The listening comprehension part of the model includes the same three processing stages posited by Anderson (1995), that is, perception, parsing, and utilization. The speech production part of the model will not be discussed here as the focus of the current study is on one-way listening.

2.2.4 Field's (2013) L2 Listening Model

One of the most recent models of L2 listening was proposed by Field (2013). Loosely drawing on Anderson's (1995) three processing stages and Culter and Clifton's (1999) model of L1 listening, Field's model classifies L2 listening comprehension into lower-level and higher-level processes. Processing at these two levels entails five operations: (a) input encoding, (b) lexical search, (c) parsing, (d) meaning construction, and (e) discourse construction. The first three operations are seen as constituting lower-level processes, which take place when a message is being decoded into language, while the latter two are included in higher-level processes associated with meaning construction.

Lower-level processes start with input decoding, involving the process of converting a sequence of acoustic signals into abstract representations that match the phonological system of the target language. This process requires decoding at the phoneme and syllable levels. The process of lexical search concerns identifying lexical items that best correspond to spoken word forms. In the process of parsing, listeners separate units in connected speech and construct propositions with the help of their understanding of standard word order, intonation group boundaries, as well as their syntactic knowledge. At the first stage of the higher-level processing, that is, meaning construction, listeners start to relate the proposition obtained from lower-level processing to their own schemata or concepts of knowledge that they have developed in order to build the actual meaning of what they have heard. The other higher-level processing stage, that is, discourse construction, includes four sub-processes that learners apply to construct the meaning of a spoken text: selecting, integrating, self-monitoring, and structure-building. Selection assists learners to notice information essential to the topic or the speaker's goal. Integrating entails adding a new piece of information to the discourse representation being developed. Self-monitoring enables learners to evaluate this information for consistency with what has been processed before. Structure-building is prioritizing and organizing stored information according to its level of relevance and importance.

2.2.5 Summary of the Cognitive Models of L2 Listening

Although these four models use different terms to describe listening processes, they all describe listening comprehension as involving complex cognitive processing that occurs interactively. Listeners engage in these processes in a parallel rather than a linear fashion, drawing on various types of knowledge sources, such as linguistic knowledge (i.e., phonological, lexical, and syntactic knowledge), external knowledge (i.e., world knowledge and knowledge of the speaker and the situation), and pragmatic knowledge (i.e., knowledge of the relationship between linguistic form and speaker intentions). To be more specific, the cognitive listening processes in Vandergrift and Goh's (2012) model are congruent with those in Field's (2013) model, as both models were developed based on Anderson's (1995) three processing stages. In contrast, Rost's (2011) model exclusively emphasizes neurological processing as the basic processing layer underlying the rest of the layers (i.e., linguistic, semantic, and pragmatic processing). Apart from this, the linguistic, semantic, and pragmatic processing layers in Rost's (2011) model seem to overlap with the processing stages described by Vandergrift and Goh (2012) and Field (2013). Considering that the four models are fundamentally similar, Field's (2013) model, which has also been employed by previous research on test-takers' listening processes (e.g., Holzknecht, 2019; Holzknecht et al., 2017), was used in the present study to analyze learners' listening processes.

2.2.6 Listening Strategies

The processing of L2 oral input, as described in the cognitive models of listening, is restricted by two factors: the amount of knowledge that listeners possess, both linguistic and non-linguistic knowledge, and the level of expertise in language processing (Field, 2013). Due to limitations in knowledge and expertise, L2 learners often have to rely on listening strategies to facilitate listening processes and fill gaps in understanding. Even competent learners may use some strategies to solve comprehension problems, for example, in conditions of noise or topic unfamiliarity (Field, 2013). Hence, the effective use of strategies should be considered an important part of L2 listening proficiency (Field, 2013; Vandergrift & Goh, 2012).

Drawing on O'Malley and Chamot's (1990) work on L2 learning strategies, Vandergrift and Goh (2012) created a taxonomy of listening strategies, including the following strategies:

1. Planning: Developing awareness of what needs to be done to accomplish a listening task, developing an appropriate action plan and/or appropriate contingency plans to overcome difficulties that may interfere with successful completion of a task.
2. Focusing attention: Avoiding distractions and heeding the auditory input in different ways, or keeping to a plan for listening development.
3. Monitoring: Checking, verifying, or correcting one's comprehension or

performance in the course of a task.

4. Evaluation: Checking the outcomes of listening comprehension or a listening plan against an internal or an external measure of completeness, reasonableness, and accuracy.

5. Inferencing: Using information within the text or conversational context to guess the meanings of unfamiliar language items associated with a listening task, to predict content and outcomes, or to fill in missing information.

6. Elaboration: Using prior knowledge from outside the text or conversational context and relating it to knowledge gained from the text or conversation in order to embellish one's interpretation of the text.

7. Prediction: Anticipating the contents and the message of what one is going to hear.

8. Contextualization: Placing what is heard in a specific context in order to prepare for listening or assist comprehension.

9. Reorganizing: Transferring what one has processed into forms that help understanding, storage, and retrieval.

10. Using linguistic and learning resources: Relying on one's knowledge of the first language or additional languages to make sense of what is heard, or consulting learning resources after listening.

11. Cooperation: Working with others to get help on improving comprehension, language use, and learning.

12. Managing emotions: Keeping track of one's feelings and not allowing negative ones to influence attitudes and behaviors (pp. 277–284).

These strategies can be categorized into (a) cognitive (inference, elaboration, prediction, contextualization, reorganizing, and using linguistic and learning resources),

(b) metacognitive (planning, focusing attention, monitoring, and evaluation) and (c) socio-affective (cooperation and managing emotions). In addition to the taxonomy proposed by Vandergrift and Goh (2012), Field (2008) and Rost (2011) added compensatory strategies that help listeners deal with breakdowns in communication. Field (2008) distinguished between four major types of compensatory strategies:

1. Avoidance strategies. Learner gets by without the missing or uncertain piece of input.

2. Achievement strategies. Learner attempts to make maximum sense of what has been decoded.

3. Repair strategies. Learner appeals for help.

4. Pro-active strategies. Learner plans her behaviour in a way that might enable problems of understanding to be avoided (p. 298).

Rost (2011) also listed several commonly noted compensation strategies to help overcome communication breakdowns and maintain effective communication:

1. Skipping: omitting a part or a block of text from processing for comprehension.

2. Approximation: using a superordinate concept that is likely to cover the essence of what has not been comprehended; constructing a less precise meaning for a word or concept than the speaker may have intended.

3. Filtering: compressing a longer message or set of propositions into a more concise one. (This is different from skipping or approximation, which are 'reduction' strategies, because filtering involves active construction of a larger semantic context.)

4. Incompletion: maintaining an incomplete proposition in memory, waiting until clarification can be obtained.

5. Substitution: substituting a word or concept or proposition for one that is not

understandable (p. 62).

Given that strategies are frequently used by listeners to construct meaning (Rost, 2011) for various reasons (e.g., natural limitations of memory, topic unfamiliarity, and the presence of distractions), the use of strategies is considered an indispensable aspect of listening comprehension in the present study.

2.2.7 Summary

To summarize, cognitive models of L2 listening generally view listening comprehension as a result of cognitive processes and listening strategies working in parallel and interactive ways. Although some of the models (e.g., Field, 2013; Rost, 2011) acknowledge the importance of visual information in constructing meaning from auditory input, none of them explicitly explains the processes involved in integrating aural and visual information. Theoretical work on how learners process multimodal L2 input involving listening is still limited. Therefore, the next section presents an overview of models of multimedia learning put forward in cognitive psychology to enable conceptualizing L2 multimodal processing and learning, the focus of this research.

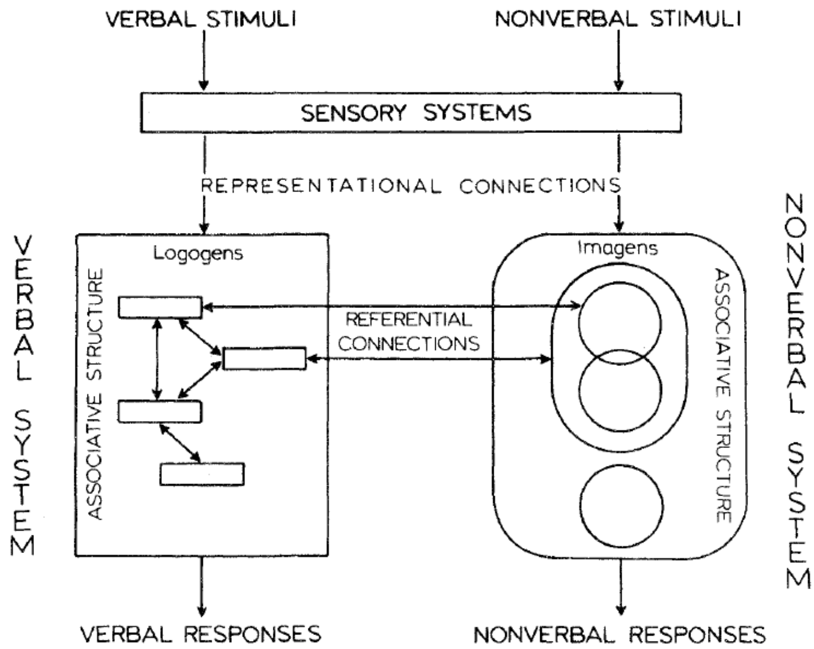
2.3 Multimodal Processing

In cognitive psychology, the term multimedia refers to the use of words and pictures to present information together (Mayer, 2014b). Words consist of verbal information presented in written or spoken form, whereas pictures include visual pictures (e.g., diagrams, photos, and animations) and auditory pictures (e.g., the call of a bird) (Schnotz, 2014). Multimedia learning, therefore, refers to situations in which people learn from words and pictures (Mayer, 2014b). In an attempt to shed light on how learners construct meaning from words and pictures, various models of multimodal input processing have been proposed in the field of cognitive psychology, with the following being particularly influential: Paivio’s (1986) dual coding theory, Mayer’s (2014c) cognitive theory of multimedia learning, Schnotz's (2014) integrative model of text and picture comprehension, and Hegarty's (2014) information processing model for understanding physical systems. To identify a model that can inform our understanding of multimodal processing in an L2 setting, these models are described in detail below.

2.3.1 Dual Coding Theory

Paivio's (1986, 2014) dual coding theory assumes that human information-processing system is composed of two cognitive subsystems, a verbal system for dealing with language and a nonverbal system for dealing with nonlinguistic objects and events (see Figure 2). The two systems are functionally independent yet interconnected. Verbal and nonverbal information is processed differently in each system, creating separate representations which Paivio (1986 p. 54) termed "logogens" (i.e., spoken or written verbal entities organized in terms of associations and hierarchies) and "imagens" (i.e., mental images and other non-verbal entities organized in terms of part-whole relationships). The two systems are connected through referential connections. Representation in the verbal system can be activated by those in the nonverbal system or vice versa. For example, a particular word in the verbal system can evoke related images in the nonverbal system. In addition, representations within each system are linked through associative processing. For instance, a particular word in the verbal system may elicit other related words. Likewise, a particular image can elicit related images in the nonverbal system. Based on this assumption, Paivio (2014) argued that information encoded both verbally and nonverbally should be better comprehended and recalled than information encoded in only one system.

Figure 2 Dual Coding Theory (Paivio, 1986, p. 67)

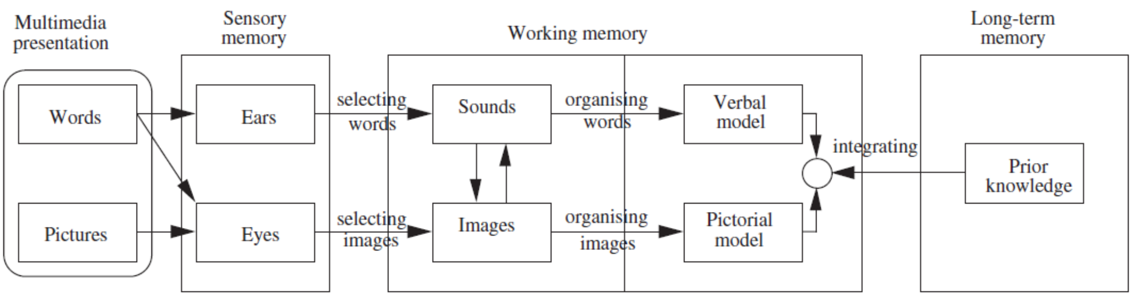


2.3.2 Cognitive Theory of Multimedia Learning

Partially inspired by Paivio’s (1986) dual coding theory, Mayer (2014c) proposed the cognitive theory of multimedia learning to explain how people learn from multimedia presentations. The theory builds on three cognitive assumptions: (a) humans possess separate information processing channels for auditory/verbal and visual/non-verbal information (Baddeley, 1992; Paivio, 1986); (b) there is only a limited amount of processing capacity available within each channel at one time (Baddeley, 1992); and (c) learning occurs when humans engage in active cognitive processing, including attending to relevant information, constructing selected information into coherent mental representations, and integrating the mental representations with existing knowledge (Wittrock, 1989).

According to the cognitive theory of multimedia learning, human information-processing system contains three memory stores (i.e., sensory memory, working memory, and long-term memory) and two channels (i.e., verbal channel and visual channel). As shown in Figure 3, words and images come in from the external environment as multimedia presentations and then enter sensory memory through eyes and ears. Sensory memory holds incoming information for a very brief time. If learners pay attention to the spoken words and pictures, it moves to working memory as sounds and images, respectively. The sounds can be mentally converted to images for processing and vice versa. Then, the learners mentally organize the words and images into coherent cognitive representations, namely, verbal and pictorial models in working memory. Finally, learners connect the two models together with relevant prior knowledge stored in long-term memory. In short, the theory predicates that learners need to engage in five cognitive processes for learning to take place in multimedia environments: (a) selecting relevant words, (b) selecting relevant images, (c) organizing the selected words into a verbal representation, (d) organizing selected pictures into a pictorial representation, and (e) integrating the verbal and pictorial representations with existing knowledge.

Figure 3 Cognitive Theory of Multimedia Learning (Mayer, 2014c, p. 52)

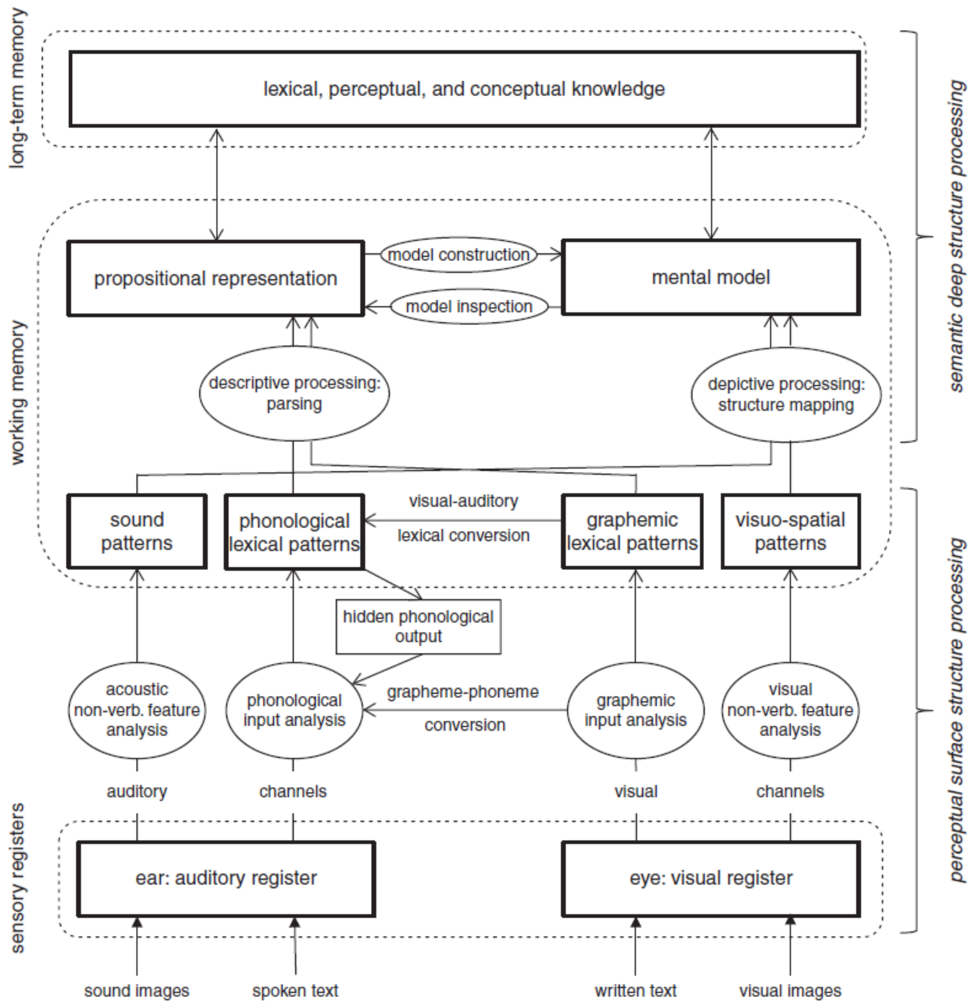


2.3.3 Schnotz's (2014) Integrative Model of Text and Picture Comprehension

Drawing on the assumptions of dual-coding (Paivio, 1986) and limited capacity (Baddeley, 1986, 2000), Schnotz's (2014) integrative model of text and picture comprehension emphasizes the representational differences between texts and pictures. As depicted in Figure 4, the model assumes that there are two levels of processing:

perceptual surface structure processing (bottom-up processing) and semantic deep structure processing (top-down processing). Perceptual processing takes place in visual and auditory channels where verbal information (spoken or written) and pictorial information (visual or auditory) is transmitted to working memory, respectively. Semantic processing, on the other hand, occurs in working memory. A descriptive and a depictive subsystem are responsible for processing verbal and pictorial information, respectively. Based on different knowledge sources, descriptive processing leads to propositional representations which then form a mental model (i.e., representations of the key parts of the presented material and their relations) through model construction. Depictive processing, on the other hand, directly leads to the mental model.

Figure 4 Integrative Model of Text and Picture Comprehension (Schnotz, 2014, p. 83)



Specifically, Schnotz's model consists of four main parts: listening, reading, visual picture, and auditory picture comprehension. In the case of listening comprehension, auditory-verbal information first enters the auditory register through the ear. Learners then recognize phonological lexical patterns via phonological input analysis (i.e., identifying phonemes in the acoustic input and forming phonological lexical patterns). This process seems to be congruent with the stages of input decoding and lexical search in Field's (2013) model. Further descriptive processing (i.e., parsing of word sequences and further semantic analysis) of phonological lexical patterns results in propositional representations of the semantic content, which is similar to the stages of parsing and meaning construction in Field's model. In terms of reading comprehension, visually presented verbal information first enters the visual register through the eye and is then subjected to graphemic input analysis (i.e., identifying graphemes in the visual input and recognizing graphemic lexical patterns). Next, propositional representations are formed via descriptive processing of graphemic lexical patterns. Guided by learners' cognitive schemata, the propositional representations are then transformed into a mental model through model construction (i.e., connecting propositional representations with structural characteristics of the mental model). This step broadly corresponds to the processing stage of discourse construction in Field's (2013) model.

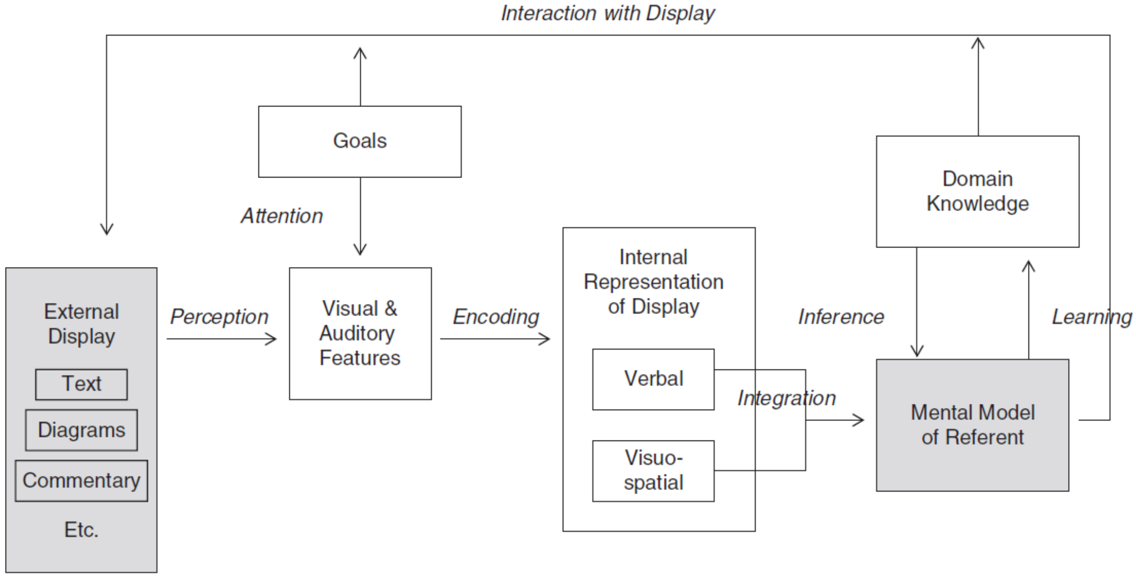
As for visual picture comprehension, visual pictorial information first enters the visual register through the eye. Learners form visuospatial structures via visual non-verbal feature analysis, that is, identifying and discriminating graphic entities in the visual pictorial input and organizing these entities according to Gestalt laws (i.e., a set of principles that describe how people visually perceive and organize complex images, such as the law of proximity which suggests that objects that are close to each other are perceived as belonging together), resulting in visuospatial patterns in working memory as visual perceptual representations. This is followed by depictive processing, that is, mapping visual spatial relations onto mental semantic relations, guided by learners' schemata and resulting in a mental model. The model can then be used by model inspection for reading new information and elaborating the evolving propositional representations. Auditory picture comprehension, the last component in Schnotz's model, is not directly related to the current study and therefore is not discussed here.

2.3.4 Hegarty's (2014) Information Processing Model

Hegarty's (2014) information processing model has a particular focus on how people understand and learn physical systems, that is, how people construct an internal presentation of physical systems from external displays. According to Hegarty (2014), physical systems include (a) man-made mechanical systems (e.g., pulley systems), (b) biological, chemical, and physical systems (e.g., organs and organisms), and (c) atmospheric processes (e.g., weather maps). External displays contain verbal materials (e.g., texts and commentaries) and/or visual-spatial materials (e.g., diagrams and animations). The internal presentation, also called a mental model, is defined as "a characterization of the knowledge and cognitive processes that enable humans to understand and predict the behavior of a physical system" (p. 674).

Figure 5 represents the cognitive processes involved in understanding physical systems. The information in the display is first perceived by visual or auditory senses. What is encoded in this process depends on learners' attention, which might be driven by the learners' goals or attracted by salient parts in the display. These basic perceptual, attentional, and encoding processes result in a partially constructed mental model. To complete the model, learners need to integrate the information presented in different modalities. This might involve translating verbal information to visual-spatial information or vice versa, as well as making some inferences from the information presented in the display based on their prior knowledge or other inference mechanisms (e.g., decoding diagrammatic conventions).

Figure 5 Information Processing Model for Understanding Physical Systems (Hegarty, 2014, p. 675)



Hegarty (2014) also introduced a distinction between two major types of external displays used in science education to represent physical systems: diagrams and verbal materials. Diagrams can be classified as static diagrams and animated diagrams. Static diagrams are “visual-spatial arrays in which information is communicated by spatial properties such as shape, location, and adjacency of parts” (p. 678). There are three important characteristics of static diagrams: (a) they are iconic in the sense that they represent spatial properties of their referents; (b) they are semantic as they do not necessarily represent all of the visual features of the referents; and (c) they can visualize information that is not visible in reality. Hence, understanding static diagrams often depends on conventions, such as arrows (i.e., objects used to draw attention or to show directions or processes), naming labels (i.e., labels providing the names of concepts or objects), and coloring coding (i.e., using colors to differentiates multiple components or to emphasize a particular component) (Kottmeyer et al., 2020). Verbal materials, on the other hand, are “descriptive representations that express facts or assertions about their referents, depend on arbitrary symbols (i.e., words), and have a linear structure” (p. 681). Verbal materials take the form of either visual texts or aural commentaries. Visual texts are self-paced as they are statically presented, whereas aural commentaries are heard at a particular rate. The content of diagrams and verbal materials can be redundant (e.g., when a text describes the structure of a system depicted in a diagram) or complementary (e.g., when a text describes how parts of a system function while a diagram demonstrates the structure of a system).

2.3.5 Summary of the Multimodal Processing Models

Now I turn to a discussion of models of multimedia learning with a view to identifying a multimodal processing model that can serve as the framework for analyzing L2 learners’ multimodal listening processes and learning behavior for the purposes of this research. First, I will highlight the similarities of the frameworks. Building on Paivio’s (1986) dual coding theory, the models developed by Mayer (2014c), Schnotz (2014), and Hegarty (2014) assume that verbal and pictorial information is processed in different systems, and multimedia learning overall relies on the activation of several cognitive processes, such as selection of verbal and graphical information, organization of the selected information for further processing, and integration of information from different sources. The three models also share the view that these cognitive processes do not necessarily occur in a linear fashion but in an interactive manner, and the mental model developed during multimodal processing is informed by prior knowledge that is not presented explicitly in the external materials. Lastly, all three models focus on multimodal learning from instructional pictures, as opposed to decorative pictures that are designed primarily for aesthetic appeal. The rationale behind this preference lies in the fact that decorative pictures provide minimal information about learning contents and thus do not directly contribute to the construction of mental models (Schnotz, 2014).

Turning to differences, Mayer assumed that information is processed in both channels (i.e., verbal channel and visual channel) independently until the verbal and pictorial mental models are established. These two mental models are subsequently integrated with the help of prior knowledge retrieved from long-term memory. Schnotz and Hegarty, however, posited that information processing results in only one modality-unspecific mental model constructed from elements of pictures and words. Furthermore, while Mayer and Hegarty assumed symmetric comprehension between verbal and pictorial processes, Schnotz hypothesized that picture comprehension provides more direct access to mental model construction than text comprehension because spatial relations from the picture can be directly mapped onto conceptual relations in the mental model. Additionally, Schnotz's model distinguished the processing of auditorily presented words from visually presented words. Considering that the viewing materials used in the current study consist of a combination of auditory, written, and pictorial information, Schnotz's model was adopted with the aim of providing a more comprehensive analysis of learners' multimodal processes. The present study also examined vocabulary learning from multimodal input, in which vocabulary was presented through aural commentaries and instructional diagrams complementing each other, corresponding to Hegarty's classification of external displays. Thus, Hegarty's model formed the basis for categorizing input sources making learners aware of lexical items.

2.3.6 The Role of Social Cues in Multimedia Learning

While models of multimodal processing allow for conceptualizing learning through verbal materials and instructional pictures, they need to be complemented with theories explaining social cues (e.g., a speaker's eye contact, facial expressions, movements, and gestures) when studying L2 viewing, given that speakers' social cues are often provided alongside instructional pictures to foster learning. It was hypothesized that social cues in multimedia instructional input can generate a feeling of social presence (i.e., a feeling of interacting with another social being), motivating learners to engage more deeply in learning processes (Mayer, 2014a). Moreover, social cues may direct learners' attention to the information that a speaker is referring to (e.g., pointing cues), which is particularly crucial for the first step in multimedia learning according to Mayer's (2014c) cognitive theory of multimedia learning, namely, selecting relevant verbal and pictorial information, as information needs to be attended to in order to be available in working memory (van Gog, 2014). To be more specific, what information will be attended to is partly determined by the characteristics of the learning material. Students with limited prior knowledge of a task, for example, may likely be attracted to perceptually salient features (Lowe, 1999, 2003), even though these features may not always be the most pertinent to the given task. When the learning materials are dynamic and transient (e.g., video lectures), attending to salient but less relevant information means that relevant information may not be attended to promptly, and so may no longer be available for further processing (de Koning et al., 2009). Therefore, instructors' signaling gestures can guide learners' attention to the essential elements of multimodal materials and facilitate subsequent learning.

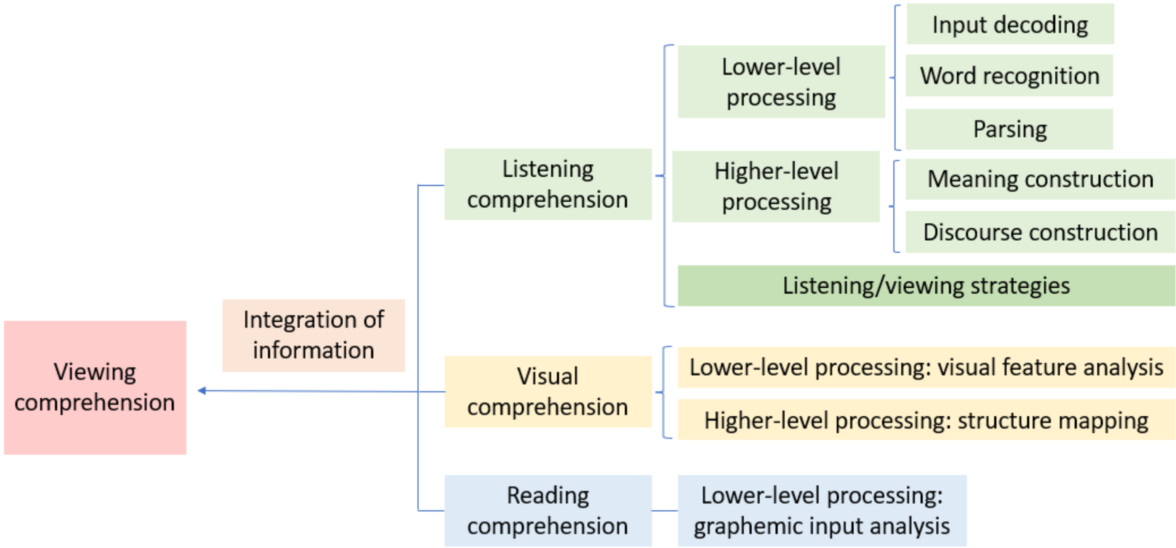
Despite having the potential to effectively promote engagement and guide learners' attention, it is argued that the instructor's presence may divert learners' attention away from relevant information and subsequently hamper learning outcomes (Colliot & Jamet, 2018). This is because different types of visual stimuli (e.g., written words, instructional pictures, and social cues) are all processed in a visual channel with a limited capacity, and learners have to split their attention among the visual stimuli (Mayer, 2015c). As a result, fewer attentional resources would be available to process information that may lead to acquisition of knowledge.

2.3.7 The Integrated Model Used in the Present Study

Drawing on previous research on L2 listening processes and multimedia learning, I constructed an integrated model of multimodal comprehension for the purposes of this study. Figure 6 summarizes five major components in the model, including listening comprehension, listening/viewing strategies, visual comprehension, reading comprehension, and integration of information. First, based on Field’s (2013) cognitive processing model, listening comprehension is subdivided into lower-level and higher-level processes. The lower-level processes contain input decoding, word recognition, and parsing, whereas the higher-level processes consist of meaning construction and discourse construction. The model also includes Vandergrift and Goh’s (2012) taxonomy of listening strategies. Some cognitive strategies in the taxonomy were dropped as they are not applicable to the viewing task used in the present study, such as cooperation and using external learning resources. In addition, the strategy of inferencing was further categorized as linguistic (i.e., using information in the text to guess the meaning of unfamiliar words) and kinesic inferencing (i.e., using facial expressions, body movements, and gestures to guess the meaning of unknown words), addressing learners’ viewing strategies of using social cues to aid comprehension. Compensation strategies proposed by Field (2008) and Rost (2011) were also added to the model.

Following Schnotz’s (2014) integrative model of text and picture comprehension, visual picture comprehension includes visual feature analysis (lower level) and depictive processing (higher level). Reading comprehension only involves lower-level processing, namely, graphemic input analysis, as the video lecture used in the study only contains individual words. The last component in the model is the integration of information from different sources to construct a coherent mental model, which involves lower- and higher-level of processing: linking individual words and graphical elements, and connecting propositional descriptions with structural characteristics of the mental model.

Figure 6 An Integrated Model of Multimodal Comprehension Developed for the Current Study



2.3.8 Summary

To summarize, this chapter provides an overview of models that aim to explain the comprehension of multimodal input drawing on perspectives of cognitive psychology. The theoretical frameworks propose that multimedia comprehension involves active knowledge construction, rather than passive internalization of learning materials. This relies on a series of cognitive processes that occur in an interactive manner. It also emerged from the review that multimedia learning environments have the potential to facilitate learners' development of linguistic knowledge through exposure to rich input. Nonetheless, scant attention has been paid to the contribution of multimodal instructional materials to language acquisition and more specifically to vocabulary acquisition. This is an important area for research, given that video-based lectures have become an essential component of university education and a major source of input for students in L2-medium education to acquire specialized vocabulary in their field of study. Additionally, specialized vocabulary is closely related to the academic success of L2 learners/users (Dang, 2020), owing to its unique feature of representing complex concepts in specific disciplines (Nation, 2013; Schmitt, 2010). In light of this, one of the focal interests of the present study was to fill this research gap. Now I turn to a review of previous literature relevant to the acquisition of technical vocabulary from instructional materials.

2.4 Acquiring Technical Vocabulary from L2 Input

In academic settings, students who attend L2-medium universities repeatedly encounter specialized terms of their field of study, as well as a large number of nonspecialized infrequent words (Parry, 1991). Consequently, they may incidentally "pick up" new words related to their field of study while attending to the overall meaning of academic input. This process is referred to as incidental vocabulary learning, generally defined as learning of vocabulary knowledge that occurs as a by-product of engaging in a meaning-focused task, such as reading novels and watching movies (Ellis, 1999). From a methodological perspective, incidental vocabulary learning has also been operationalized as acquisition of L2 vocabulary under conditions when learners are not informed of upcoming vocabulary post-tests (Hulstijn, 2003). The by-product definition has been predominantly used in incidental vocabulary learning research and was adopted in the present study. Intentional learning, on the other hand, entails engaging in activities where learning is deliberate and goal-directed (Wode, 1999).

Although researchers have found that intentional learning leads to better vocabulary acquisition (Laufer, 2003), relying solely on deliberate learning is inefficient for a number of reasons. First, the amount of vocabulary that can be explicitly taught and learned in a classroom setting is inherently limited (Webb & Nation, 2017). Also, specialized vocabulary often requires an in-depth understanding of a specific discipline to be effectively taught (Coxhead, 2018). As a result, language teachers who lack sufficient background knowledge of the discipline may struggle to teach such vocabulary. Additionally, vocabulary development is an incremental process, with new words requiring multiple encounters before they can be acquired and integrated into learners' mental lexicon (Nation, 2013; Webb & Nation, 2017). Therefore, apart from explicit vocabulary instruction, incidental vocabulary acquisition through exposure to the target language during academic learning is also an important means of developing specialized vocabulary knowledge.

2.4.1 Defining Technical Vocabulary

Learning technical vocabulary, a type of specialized vocabulary, from academic input is a specific form of incidental vocabulary learning (Gablasova, 2014), given that it is often acquired automatically as a consequence of learning new subject knowledge (Schmitt, 2010). Technical vocabulary has been broadly defined as words that are closely related in meaning to the content of a specific subject (Nation, 2013), ranging from words that are exclusively used within a subject (e.g., *pericardium* in anatomy) to high-frequency words in general language that may or may not have an additional subject-specific meaning (e.g., *memory* in computer science and *brain* in neuroscience) (Liu & Lei, 2020; Nation, 2013). The current study only focused on subject-specific words, meaning that acquiring new technical vocabulary involves simultaneous learning of new lexical forms as well as new concepts (Schmitt, 2010). For example, an L2 user may know nothing about neuron structures, thus may be learning the form and meaning of the term *soma* through listening to its explanation "soma, or cell body, is the core section of the neuron, which contains genetic information and provides energy for the neuron". This is considered fundamentally different from much of incidental vocabulary acquisition that involves attaching new L2 forms to already known L1 concepts (Schmitt, 2010).

2.4.2 Acquiring Technical Vocabulary from Context

As demonstrated in the example presented above, a crucial factor for learners to incidentally acquire technical vocabulary is the presence of rich contextual information. In textbooks and lectures, definitions are often used to provide an appropriate amount of information from which the meaning of technical words can be inferred (Flowerdew, 1992; Lessard-Clouston, 2009). Chaudron (1982) was among the first to research spoken definitions in lectures, or what he called "elaboration" (p. 171). He classified elaboration as either explicit (defining, questioning, naming, and describing) or implicit (paraphrasing, parallelism, and apposition). Flowerdew (1992) later proposed a more fine-grained classification of spoken definitions, including the categories of formal definition (i.e., a precise statement of a word), semi-formal definition (i.e., a statement that only identifies key characteristics of a word), and substitution (i.e., words or phrases that is substituted for the target word, including synonym, paraphrase, and derivation).

Having recognized the importance of context in facilitating language acquisition, a Content and Language Integrated Learning (CLIL) approach has attracted much interest. It aims to promote both content and language learning by delivering a subject in the target language (Coyle et al., 2010). This dual purpose is what mainly distinguishes CLIL from other bilingual programs whose primary focus is on developing students' general L2 proficiency. CLIL does not represent a completely novel concept; instead, it is posited to be a synthesis of various theories and approaches (Mehisto et al., 2008). It is considered the most recent developmental stage within the communicative language teaching (CLT) framework, given that it creates a context to facilitate authentic and meaningful communication while simultaneously affording students repeated exposures to the target language (Coyle et al., 2010). Additionally, CLIL incorporates key characteristics of TBLT as students engage in real-life tasks where their focus is on a given task rather than the language itself (Georgiou, 2012), such as doing a collaborative experiment using the target language. Although CLIL is closely connected with TBLT, the present study centers on the latter approach. A comprehensive discussion of TBLT will be presented in Section 2.5.

2.4.3 Measuring Vocabulary Knowledge

Following the discussion on how technical vocabulary can be acquired from context, it is worth turning our attention to different methods of assessing the efficacy of this acquisition process. Research has proposed different ways to measure the depth of vocabulary knowledge (i.e., how well a word is acquired), which can be broadly categorized into three approaches: a developmental approach, a lexical network approach, and a components approach (Yanagisawa & Webb, 2019). The developmental approach conceptualizes the depth of vocabulary knowledge as development in the degree of word knowledge, from no knowledge to complete mastery. This approach involves using scales to indicate the developmental stage of lexical knowledge. The Vocabulary Knowledge Scale (VKS, Paribakht & Wesche, 1993; Wesche & Paribakht, 1996) is one of the most widely used tests adopting this approach. As presented in Figure 7, the VKS combines self-reporting and a performance test, aiming to capture the developmental stage of lexical knowledge ranging "from complete unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence" (Wesche & Paribakht, 1996, p. 29). Despite being widely used in the field of vocabulary research, the VSK has faced criticism on the basis that it sees vocabulary development as a linear process and does not provide a valid measure of different types of lexical knowledge (Schmitt 2010; Yanagisawa & Webb, 2019).

Figure 7 An example of the Vocabulary Knowledge Scale Test (Wesche & Paribakht, 1996, p. 30)

VKS elicitation scale self-report categories

Self-report categories
I. I don't remember having seen this word before.
II. I have seen this word before, but I don't know what it means.
III. I have seen this word before, and I *think* it means___________. (synonym or translation)
IV. I *know* this word. It means___________. (synonym or translation)
V. I can use this word in a sentence:___________. (If you do this section, please also do Section IV.)

The lexical network approach operationalizes depth of vocabulary knowledge as L2 learners' ability to associate different words in their mental lexicon (Yanagisawa & Webb, 2020). The free word association task has been the primary means of investigating L2 learners' lexical networks, in which learners are required to provide the first word that comes to their mind in response to a cue word (for a review, see Fitzpatrick, 2013). Building on previous research on word association, Read (1993, 1998) developed the Word Associates Format (WAF), which prompts learners to select words that are associated with the target words rather than produce them. An example of the WAF is presented in Figure 8. While the WAF is widely used to measure the depth of learners' vocabulary knowledge, a notable limitation lies in the challenges associated with interpreting and comparing test results across different studies due to the differences in selecting target words, association relationships, and test formats (Yanagisawa & Webb, 2019).

Figure 8 An example of the Word Associates Format Test (Read, 1998, p. 46)

sudden

| beautiful quick surprising thirsty | | change doctor noise school |

The third approach, the component approach (also known as the dimensions approach), involves breaking vocabulary knowledge into separate categories and quantifying learners' acquisition in regard to each dimension (Schmitt, 2010). Following this approach, Nation (2001) proposed the most well-known and comprehensive specification of work knowledge aspects (see Table 1). The model includes three broad aspects of knowing a word: form (spoken form, written form, and word parts), meaning (form and meaning, concept and referent, and associations), and use (grammatical functions, collocations, constraints on use). One major advantage of using the component approach is its high construct validity, as different aspects of word knowledge can be investigated independently (Schmitt, 2010; Yanagisawa & Webb, 2019). Webb (2007) is the most striking example, in which ten different tests were used to measure five dimensions of vocabulary knowledge (i.e., orthography, syntax, grammatical, functions, association, and meaning-form connections) at both levels of recognition and recall. However, using this approach can be time-consuming, and the number of words that can be assessed in each individual study is limited (Schmitt, 2010). In light of the comprehensive nature of this approach, it is often unfeasible and unnecessary to measure all aspects of lexical knowledge, thus researchers should decide which aspects to assess based on their specific research aims (Yanagisawa & Webb, 2019).

Table 1 Model of Aspects of Vocabulary Knowledge (Nation, 2001, p. 27)

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use	R	Where, when, and how often would we expect to meet this word?
	(register, frequency ...)	P	Where, when, and how often can we use this word?

Note: R = receptive knowledge, P = productive knowledge

Among various aspects of lexical knowledge, form-meaning association is one of the most commonly measured lexical aspects in L2 vocabulary research, as it is the first and most important aspect which learners must acquire (Laufer & Goldstein, 2004; Schmitt, 2010). In Nation's (2001) model, the form and meaning aspects are further distinguished as receptive and productive. Receptive knowledge refers to "perceiving the form of a word while listening or reading and retrieving its meaning", whereas productive knowledge involves "wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form" (Nation, 2013, p. 47). To avoid inconsistency in measuring receptive and productive knowledge, Laufer and Goldstein (2004, p. 405) proposed four degrees of vocabulary knowledge (see Table 2) based on learners' competence in "supplying the form for a given meaning versus supplying the meaning for a given form" and "being able to recall versus only being able to recognize (whether form or meaning)."

Table 2 Degree of Vocabulary Knowledge (Laufer & Goldstein, 2004, p. 407)

	Recall	Recognition
Active (retrieval of form)	Supply the L2 word	Select the L2 word
Passive (retrieval of meaning)	Supply the L1 word	Select the L1 word

To make Laufer and Goldstein's (2004) original terms "active" and passive" more transparent when assessing the depth of vocabulary knowledge, Schmitt (2010, p. 86) relabeled the four categories as "form recognition", "form recall", "meaning recognition", and "meaning recall" (see Table 3). Form recognition is when the meaning is presented and the form needs to be identified. Form recall refers to the case when the meaning is given and the form must be produced. Meaning recognition is when the form is given, and the meaning needs to be recognized. Meaning recall means the form is given, and the meaning needs to be produced. The use of these four form-meaning link categories is assumed to help gain a fuller picture of the incremental nature of the vocabulary learning processes and reveal learners' degree of mastery (Schmitt, 2010).

Table 3 Degree of Vocabulary Knowledge (Schmitt, 2010, p. 86)

Word knowledge	**Word-knowledge tested**	
Given	*Recall*	*Recognition*
Meaning	Form recall (supply the L2 item)	Form recognition (select the L2 item)
Form	Meaning recall (supply definition/L1 translation, etc.)	Meaning recognition (select definition/L1 translation, etc.)

2.4.4 Empirical Studies on Incidental Acquisition of Technical Vocabulary

To date, few studies have used different measures to gauge learners' lexical knowledge gained from academic viewing. Smidt and Hegelheimer's (2004) study is one of the earliest published studies that addressed the issue. Twenty-four English as a Second Language (ESL) students were required to complete a computer-assisted language learning (CALL) activity, that is, watching a short video of an academic lecture on horticulture with an online dictionary. The lecture contained auditory as well as different types of visual input (i.e., a lecturer's presence, instructional pictures, and keyword notes). The participants' knowledge of 20 words/phrases selected from the lecture was pre- and post-tested using partial dictation tasks. The results suggested that performing the CALL activity enhanced the participants' vocabulary acquisition. It is not clear, however, whether the learning gains were from the lecture itself or from consulting the online dictionary.

Vidal's (2003) experiment focused exclusively on the role of academic lectures, in which 116 English as a Foreign Language (EFL) students were asked to watch three videotaped lectures on tourism with 36 target words embedded. The words were categorized as technical, academic, and low-frequency. Technical words were defined as words that are "necessary for the development of the topic of the lecture" (p. 62). Learners' knowledge of target items was measured in a pre-test, immediate post-test, and delayed post-test, respectively, using an adapted version of the VKS (Wesche & Paribakht, 1996). Participants' listening comprehension was assessed by true-false questions and listening cloze tests. The results demonstrated that learners made greater gains in the knowledge of technical words than academic or low-frequency words, which emphasized the importance of technical words in lecture comprehension. It was also reported that words that received more explicit elaborations achieved greater gains. Using similar research materials and instruments, Vidal (2011) compared incidental vocabulary gains from reading and listening. The results showed that both reading and watching academic lectures led to vocabulary knowledge gains, but academic reading was found to be more effective. Nonetheless, listeners made more gains on the technical items than the other two types of words, while readers only performed slightly better on the technical items.

Unlike Vidal (2003, 2011) who adopted modified reading style lectures, Yang and Sun (2013) used three authentic open-course lectures on psychology, physics, and music to investigate incidental vocabulary learning from lecture viewing with 65 ESL learners. Similarly, 33 target items were classified into technical, academic, and low-frequency words. A true-false test and the VKS (Wesche & Paribakht, 1996) were used to measure the participants' lecture comprehension and vocabulary gains, respectively. Consistent with Vidal's (2003, 2011) findings, the results indicated that learners achieved greater gains in the knowledge of technical words than the other types of words. Additionally, words with explicit elaborations were more successfully acquired than those with implicit or no elaboration. In addition, nonverbal elaboration was found to have a facilitating effect on learners' vocabulary acquisition.

Recognizing the absence of a control group in previous studies, Dang et al. (2021) looked into learning single words and collocations through lecture viewing with 55 EFL learners. They were assigned to either an experimental or a control group and were subjected to pre- and post-testing using a meaning recall and a meaning recognition test. The experimental group watched an unmodified lecture on algorithms with 50 single words and 19 collocations embedded, whereas the control group did not receive any treatment. The single words consisted of nonspecialized, administrative, academic, and technical words. The results showed that the experimental group made significant gains in the knowledge of single words at the level of meaning recall and collocations at the level of form recognition.

These studies provided evidence that learners were able to acquire vocabulary knowledge, including technical words, in the context of academic lecture viewing. However, they did not explicitly state whether learners simultaneously encountered both new L2 forms and new semantic information. In addition, none of the studies used science lectures that contain diagrams to convey information and facilitate learning. It remains unclear how learners make use of instructional pictures and verbal information to develop their knowledge of technical words during L2 lecture viewing. In terms of methodological limitations, some of the studies used VKS to measure vocabulary gains. As Schmitt (2010) pointed out, self-reported data may not be as reliable as a direct demonstration of vocabulary knowledge, and VKS does not differentiate between various aspects of vocabulary knowledge in a principled way. The component approach has been widely employed by recent empirical studies on vocabulary acquisition from L2 listening (e.g., Hatami, 2017; van Zeeland & Schmitt, 2013) and has been evident to be successful in tapping into subtle increases in vocabulary knowledge. Therefore, this approach was adopted in the current study to measure vocabulary gains.

The only existing study that has explicitly investigated the simultaneous learning of technical vocabulary and content knowledge was carried out by Gablasova (2014), with a specific focus on reading. Sixty-four ESL learners at intermediate or advanced levels were instructed to read while listening to academic texts in either their L1 or L2. Twelve technical terms new to the participants both in their L1 and L2 were chosen as target items. Participants' knowledge of the target items was measured by an immediate and a delayed meaning recall test (1 week later). The results indicated that the L2-instructed participants recalled fewer word meanings and developed a less accurate understanding of the words than the L1-instructed participants.

2.4.5 Summary

In summary, acquisition of technical vocabulary involves learning new lexical forms representing new content knowledge. This is considered more challenging than learning novel words where learners already know the underlying concepts associated with the words (Liu & Lei, 2020; Nation, 2013; Schmitt, 2010). However, little attempt has been made to investigate how technical words are processed and acquired from lecture reviewing. Considering the inherent difficulty of acquiring technical terms, it is imperative to further explore how it might be enhanced through L2 pedagogy. Among the various pedagogical options, the potential of task repetition appears to be particularly worthwhile to explore, given the proliferation of recorded lectures in academic contexts which allows students to engage in repeated listening/viewing. Besides, a few studies have found significant positive effects of task repetition on learners' vocabulary acquisition from L2 listening (e.g., Ellis & Chang, 2016; Shintani, 2012a). So far, however, little is known about whether repeated viewing of video lectures may as well facilitate L2 learners' processing and acquisition of vocabulary. To understand how best to fill this gap, the next section considers theories underpinning task repetition as well as previous research investigating its role in L2 processing and acquisition.

2.5 Task Repetition

Most previous research on task repetition has been contextualized in the context of TBLT, a pedagogical approach that has received increased research since the 1980s (e.g., Bygate et al., 2001; Ellis, 2003; Nunan, 2004). It aims to promote language learning by engaging learners in meaning-focused tasks (Ellis, 2003). A consensus on the definition of tasks, however, has not been reached. From different perspectives in research or pedagogy, there are a multiplicity of definitions proposed with each highlighting certain aspects of the construct (Bygate et al., 2001). On the basis of the diverse definitions, Ellis

(2009, p. 223) provided four defining criteria for a task: (a) “meaning” is the primary focus; (b) there is some type of “gap”, such as an information gap; (c), learners should mainly rely on their own linguistic and non-linguistic resources to communicate, and (d) there is a clearly defined outcome other than the display of correct language use.

2.5.1 Multimodal Lecture-Viewing Tasks

Although a consensus on the definition of “task” has not been reached, most definitions of pedagogic tasks either implicitly or explicitly suggest that all four language skills are involved (i.e., reading, listening, speaking, and writing) (see Table 4 for examples). Thus, it seems obvious that a task can be output-based (i.e., involving speaking or writing) and input-based (i.e., involving listening or reading). The difference between the two types of tasks lies in whether learners are requested to produce linguistic output (Ellis, 2003). Input-based tasks can be further categorized as listening input-based tasks, which often take the form of listen-and-do tasks and academic listening tasks (Ellis, 2003). Listen-and-do tasks are one-way information gap tasks that require learners to listen to instructions and perform specific actions, such as pointing to a picture to demonstrate their understanding of the commands (Shintani, 2012a, 2012b). In contrast to listen-and-do tasks which are particularly advantageous for learners in the initial stages of learning an L2, academic listening tasks are often targeting more competent L2 learners. They consist of a lecture on an academic topic, during which learners are asked to take notes (Ellis, 2003).

Table 4 Definitions of Tasks

Source	Definition
Long (1985, p. 89)	A task is “a piece of work undertaken for oneself or for others, freely or for some reward. Thus, examples of tasks include painting a fence, dressing a child, filling out a form, buying a pair of shoes, making an airline reservation, borrowing a library book, taking a driving test, typing a letter, weighing a patient, sorting letters, taking a hotel reservation, writing a cheque, finding a street destination, and helping someone across a road, in other words, by ‘task’ is meant the hundred and one things people do in everyday life, at work, at play, and in between. ‘Tasks’ are the things people will tell you they do if you ask them and they are not applied linguists.”
Prabhu (1987, p. 24)	A task is “an activity which required learners to arrive at an outcome from given information through some process of thought and which allows teachers to control and regulate that process.”
Skehan (1998, p. 95)	A task is “an activity in which: meaning is primary; there is some communication problem to solve; there is some sort of relationship to comparable real-world activities; task completion has some priority; the assessment of the task is in terms of outcome.”
Bygate et al. (2001, p. 11)	A task is “an activity which requires learners to use language, with emphasis on meaning, to attain an objective”
Ellis (2003, p. 16)	“A task is a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed. To this end, it requires them to give primary attention to meaning and to make use of their own linguistic resources, although the design of the task may predispose them to choose particular forms. A task is intended to result in language use that bears a resemblance,

Source	Definition
	direct or indirect, to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills, and also various cognitive processes."
Nunan (2004, p. 4)	A task is "a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is primarily focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form."

Ellis (2003) provided a detailed description of academic listening tasks within the TBLT framework, which consists of five major elements: goal, input, conditions, procedures, and predicted outcome. Goal refers to the general purpose of an academic listening task, such as promoting learners' ability to comprehend an academic lecture and facilitating their linguistic development. Input refers to the data provided by a task, including the lecture itself as well as other non-verbal aids (e.g., diagrams, photos, and graphs). If tasks deliver academic input in multimodal modalities, they can be further conceptualized as multimodal input-based tasks (Lee & Révész, 2020). Conditions concern how data are presented. Academic listening tasks are apparently one-way, but they can also be more colloquial and interactive. As for procedures, this covers the methodological choices available to researchers and teachers for task implementation. For instance, students can take notes as they listen to a lecture or during pauses in it. Also, learners may or may not receive guidance during note-taking.

The last component, predicted outcomes of a task, constitutes an indispensable part of the framework, given that tasks need to have an explicit and specifiable outcome to qualify as such (Ellis, 2003). The product outcome of an academic listening task can be sets of notes taken by learners, providing physical records of the lecture content from which information can be recalled and rehearsed during note-viewing (Di Vesta & Gray, 1973). In addition to this external storage function, the act of note-taking also serves as an encoding function, which may facilitate comprehension and retention of information (Di Vesta & Gray, 1973). On the other hand, the process of note-taking may impose a dual task demand on learners and potentially hinder their lecture comprehension (Chaudron et al., 1994).

2.5.2 Theoretical Underpinnings of Task Repetition

Apart from the discussion around different task types (i.e., output-based versus input-based tasks), previous TBLT research has been interested in exploring task implementation factors that might maximize the developmental benefits of engaging in task-based activities. Among various implementation factors, task repetition has emerged as a promising tool to direct learners' attention to linguistic forms during task performance. Task repetition refers to "repetitions of the same or slightly altered tasks – whether whole tasks, or parts of a task" (Bygate & Samuda, 2005, p. 43). In this conceptualization, task repetition can be generally categorized into three types: (a) exact task repetition, involving repetition of the exact same tasks; (b) procedural repetition, entailing repetition of the procedure but requiring different content knowledge; and (c) content repetition, in which the procedure changes but the content or material remains unchanged (Patanasorn, 2010).

Skehan's (1998) limited capacity model, also known as the trade-off hypothesis, provides a rationale for task repetition. It assumes that humans have limited attentional capacity and must prioritize where they allocate their attention. During initial task performance, cognitive demands on meaning would inevitably result in less attention being available to language (Skehan, 1998; Skehan & Foster, 2001). Task repetition, therefore, might be useful to ease cognitive demands as regards content and direct learners' attention to linguistic forms. In the context of L2 listening, listeners are likely to focus primarily on meaning construction during the first listening. When they have gained some familiarity with the content of the text in subsequent listening, their attentional resources that would be focused on meaning are free to be applied to the processing of language. Task repetition also enables listeners to switch attention to areas where they might have missed or had difficulties forming word-sound matches during a previous round of listening, increasing the probability that they consciously attend to new language.

Additionally, Robinson's Triadic Componential Framework (2001a, 2001b) may as well provide theoretical ground to discuss the impact of task repetition. The framework distinguished three factors that interact to influence task performance and learning, namely, task complexity, task difficulty, and task conditions. Task complexity is defined as the results of a series of information processing demands posed by the inherent characteristics of a task (Robinson, 2001b). Two cognitive factors associated with task complexity are the resource-directing and resource-depleting dimensions (Robinson, 2001a, 2001b). Raising task complexity along the resource-directing dimension draws learners' attention to the language code and "pushes" learners to use more sophisticated linguistic structures to fulfill the greater cognitive demands posed by a more complex task. For example, compared to a task with a few elements, a task with many elements requires specific lexis and complex clauses to distinguish various elements, while other factors remain constant.

In contrast, increasing task complexity along the resource-depleting dimension may lead to the distraction of attention away from linguistic aspects of a task. The resource-depleting variables that affect task complexity can be single tasks versus dual-tasks or single performance versus repeated performance. Specifically, adding a dual-task to a primary task will not direct learners' resources to certain linguistic forms, rather it will distract their attention away from the primary task. On the contrary, performing the task repeatedly may decomplexify the task, as learners gradually familiarize themselves with the given task and prepare themselves for the upcoming completion (Malicka et al., 2019).

While task complexity is mostly influenced by cognitive factors, task difficulty is primarily affected by learner factors (e.g., motivation, anxiety, and confidence) (Robinson, 2001a). Task conditions, which refer to the interactive demands of tasks, include participation variables (e.g., the direction of information flow and communication goals) and participant variables (e.g., proficiency, gender, and familiarity with task roles) (Robinson, 2001a).

2.5.3 Empirical Studies on Repetition and Listening Processes

So far, task repetition has been mostly researched in relation to output-based tasks, and a positive effect of repeated task engagement on learners' oral performance has been evidenced in many previous studies (see Bygate, 2018, for a review). Nonetheless, there have been no empirical studies investigating the effect of task repetition on learners' L2 auditory processing. In the context of assessing L2 listening, studies conducted by Field (2015) and Holzknecht (2019) have provided valuable insights into test-takers' listening processes and strategic behavior in relation to double play. It should be noted that the term *task* was used in their studies to refer to the listening text and a set of items and instructions.

Field (2015) first collected quantitative data from 73 pre-sessional students at a UK university. They took two International English Language Testing System (IELTS) listening tasks (i.e., a multiple-choice and a gap-filling task) and were made to believe that they would hear a recording once only, but were then permitted a second hearing. The results of quantitative analyses showed that test-takers' scores significantly increased after a second hearing, and the gap-filling task benefitted significantly more in terms of increased scores than the multiple-choice task. Qualitative data was then collected from verbal reports and semi-structured interviews with another 37 students. They were informed in advance that they would hear the recording twice. During each listening, the recording was paused after several questions, and the participants were asked in their L2 about how they had arrived at their answers and made use of the second play. The results of qualitative analyses indicated that, for most participants, their cognitive behavior varied considerably when allowed a second hearing. Key features of behavior changes included reduced listening anxiety and increased familiarity with the listening material, which allowed them to locate important information. In addition, although the participants' attention was still heavily focused on word-level decoding during the second listening, many of them reported that they gained a wider perspective of the listening content as well as the speaker's goals.

Holzknecht (2019) critiqued several methodological weaknesses of Field's qualitative study. First, Holzknecht pointed out that using participants' answers to the test items as stimuli during the first listening might have influenced their answer choice in subsequent listening. Pausing audio to conduct stimulated recalls might have also changed participants' natural processing of the texts. Besides, Field's stimulated recall participants were aware of the double-play condition, which could have potentially influenced their cognitive behavior compared to those who were not informed of a second hearing. Finally, participants were not given a choice between L1 and L2 to do verbal reports and interviews, although attempts were made to adjust interview questions to fit the participants' L2 proficiency level. These methodological issues were addressed in

Holzknecht's (2019) experiment, in which quantitative data were collected from 306 Austrian secondary school students. They were instructed to complete four listening tasks in either a single- or a double-play condition. The listening tasks were taken from the listening section of the standardized Austrian matriculation examination, consisting of two formats: multiple-choice and note-form (i.e., filling in gaps at the end of sentences). The participants were told from the outset that they were going to hear the recording either once or twice. They were also asked to complete questionnaires targeting their listening strategies, test-taking strategies, listening anxiety, and perceptions of the tasks after completing the test. The results of the quantitative analysis were in line with Field's findings, showing that double play significantly increased test scores.

In the second part of the experiment, Holzknecht collected qualitative data from 16 students who were asked to complete the same listening tasks under both conditions while their visual attention to the listening tasks was captured by an eye tracker. The participants were also informed of the number of times they were going to hear the recording. After the experiment, they were asked general questions about task completion and were then prompted by a recording of their eye movements overlaid with the audio of the tasks to recall their thought processes during listening. The participants could stop the recording at any time to report their thoughts, and the researcher stopped the recording when noticing unexpected eye movements and participants' reactions (e.g., laughing). All interviews were conducted in participants' L1. The stimulated recalls were qualitatively analyzed in terms of four response processes (i.e., processing, listening strategies, test-taking strategies, and anxiety levels). The results showed that double play affected all four response processes. The participants demonstrated a larger proportion of higher-level processes and a smaller number of lower-level processes in the double-play condition as compared to the single-play condition. They also engaged in a wider range of listening strategies and relied on fewer test management and test wiseness strategies. The participants' anxiety level was also significantly reduced.

2.5.4 Empirical Studies on Repetition and Vocabulary Acquisition

While previous studies have examined the impacts of repeated listening on test-takers' cognitive processes and test performance in language assessment settings, a few studies have explored acquisitional benefits of repeating listening input-based tasks within the framework of TBLT. Shintani (2012a) was one of the first to investigate the extent to which listening task repetition facilitates vocabulary learning. The participants were 30 young learners, who were assigned to a control and a task repetition group. The control group listened to songs and learned formulaic expressions over five weeks, while the experimental group completed a listen-and-do task nine times in the same period. The task involved the participants listening to instructions and pointing to a picture to demonstrate their understanding, with a view to introducing 24 preselected vocabulary items. The pre-test, post-test, and delayed post-test all took the form of communicative tasks. The task repetition group made significantly greater gains in both receptive and productive vocabulary knowledge, as reflected in their immediate and delayed post-test scores.

Ellis and Chang's (2016) study used another type of listening input-based task, namely, information transfer tasks (i.e., listening to texts and transferring information into a chart or table), to examine the effects of task repetition on L2 vocabulary learning. The participants were 130 L1 Chinese university students from three intact classes. One group heard the text once and completed the table/chart during listening, whereas the other two groups only listened to the text the first time, filled in the chart/table the second time, and revise the chart/table during the third listening. The third group also received inference training before each task. The information transfer tasks included 10 target words (five nouns and five adjectives) with each occurring twice. Learners' vocabulary gains were measured by an immediate and a delayed post-test, including a written form recognition, a meaning recognition, and a meaning recall test. The groups who repeated the listening tasks showed greater immediate gains in their receptive knowledge of the target word forms, but there was no difference among the three groups in the delayed post-test.

Although these studies shed light on the facilitating effects of repeating input-based tasks, our understanding of how repeating multimodal input-based tasks affects learners' vocabulary acquisition is still limited. To the best of my knowledge, only a study by Majuddin and Siyanova-Chanturia (2021) examined acquiring multiword expressions from repeated video viewing. The researchers recruited 122 Malaysian L2 English learners from six intact classes. They were randomly assigned to six experimental conditions, according to whether they watched a television episode once or twice with no captions, normal captions, or enhanced captions. Participants' knowledge of 18 multi-word expressions was tested using a form-recall pretest. An immediate post-test (a gap-fill transcript-based format) and a delayed post-test (consisting of a form recognition, a meaning recognition, and a meaning recall test) were also administered. The results indicated that repetition led to greater immediate recall of multi-word expressions across all caption conditions but did not emerge as a strong predictor of delayed form recall. Further research is needed to explore how repeated viewing affects the acquisition of single words.

2.5.5 Summary

To sum up, attempts have been made to investigate the effects of task repetition on learners' vocabulary acquisition from listening input-based tasks (Ellis & Chang, 2016; Shintani, 2012a), as well as test-takers' cognitive processes and strategic behavior during repeated listening in the context of L2 testing and assessment (Field, 2015;

Holzknecht, 2019). Nonetheless, it still remains unclear how repeating a multimodal input-based task affects processes underlying multimodal comprehension and acquisition of new lexis from task performance. To gain a deeper understanding of the attentional and acquisitional processes, L2 researchers have argued the need to combine multiple data sources in their investigations. Specifically, the use of eye-tracking may help gain insights into learners' visual attention to multimodal L2 input without interfering with the task at hand, whereas retrospective reports may elicit learners' conscious thought processes and the strategies they use during task performance. This triangulation approach has already been found useful in a few reading studies (e.g., Godfroid & Schmidtke, 2013; Jung & Révész, 2018), and appears promising to shed light on the complex nature of multimodal comprehension and vocabulary acquisition from viewing. The following section, therefore, provides an overview of these process measures as well as data triangulation.

2.6 Measures of Attentional and Cognitive Processes

There is a persistent methodological challenge in investigating attentional and acquisitional processes underlying L2 listening/viewing and acquisition, owing to the inherently dynamic and real-time nature of the processes. To address this challenge, researchers have turned to powerful techniques, namely, verbal reports and eye-tracking, to capture these processes.

2.6.1 Verbal Reports

Verbal reports refer to "gathering data by asking individuals to vocalize what is going through their minds as they are solving a problem or performing a task" (Gass & Mackey, 2016, p. 13). Different forms of verbal reports can be roughly categorized into either concurrent or retrospective based on the time of reporting (Bowles, 2008, 2010; Egi, 2004; Ericsson & Simon, 1993). Concurrent reports are collected as learners verbalize their thoughts while simultaneously performing a task, such as think-aloud and note-taking, whereas retrospective reports are collected after completing a task, often taking the form of questionnaires, diaries, and stimulated recalls (Bowles, 2008, 2010; Egi, 2004). Although verbal reporting has been widely used in L2 research to gain access to learners' cognitive processes that are not available by other means, there are some constraints associated with this technique, often discussed with respect to reactivity and veridicality. Reactivity concerns the danger that verbalization during task performance could alter cognitive processing (Russo et al., 1989), and veridicality involves the accuracy of verbal reports (Bowles & Leow, 2005; Egi, 2004, 2008). Concurrent verbal reports might potentially have a reactive effect on learning processes, given that verbalizing while simultaneously performing a task may force learners to engage in dual-task performance.

On the other hand, the danger of learners reporting inaccurately is higher for retrospective rather than concurrent reports because of the time lag between the task and the retrospective report. This potential threat of veridicality, however, can be mitigated in part by a presentation of recall prompts, such as audio or video recordings of learners' performance, to support the recall of information (e.g., Adams, 2003; Egi, 2004, 2008; Gass & Mackey, 2016). This specific form of retrospective report is known as stimulated recall, characterized by the use of stimuli to activate and refresh recollections of learners' cognitive processes so that they can be accurately recalled (Gass & Mackey, 2016). Stimulated recalls are particularly useful in terms of researching listening processes, as learners cannot think aloud while listening to long texts. The risk of veridicality could also be lessened by reducing the time delay between the task and subsequent recall (Ericsson & Simon, 1987, 1993). To further increase the validity of recall protocols, transcripts need to be coded following a coding scheme (Kasper, 1998), and by more than one researcher to check coder-reliability.

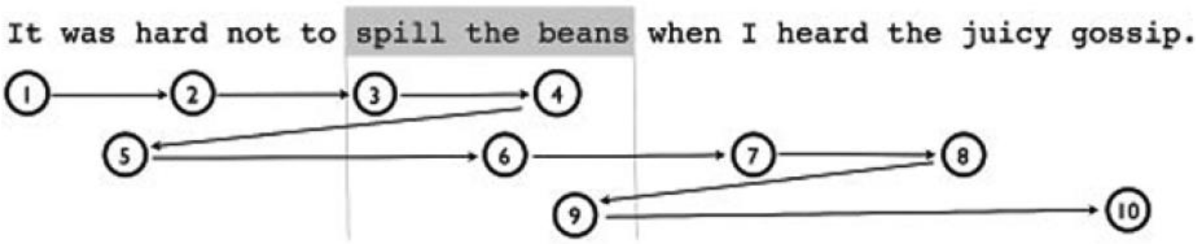
In recognition of its benefits, stimulated recall has been used in a few studies investigating test-takers' listening processes (e.g., Badger & Yan, 2012; Field, 2013, 2015). Written items and audio recordings of texts were often used in the studies to elicit retrospection. More recent research has used recordings of test-takers' eye movements during listening as stimuli (e.g., Holzknecht, 2019; Holzknecht et al., 2017; Winke & Lim 2014), which was assumed to help participants provide a more comprehensive and objective report of their thought processes during task completion (van Gog et al., 2009). The stimulated recall data collected in this manner could also be triangulated with the interpretations made on the basis of eye-movement data analyses, as eye-tracking taps into learners' real-time viewing behavior, whereas stimulated recall can provide further information about listeners' conscious cognitive activities during viewing.

2.6.2 Eye-tracking

Eye-tracking has been widely used in L2 research because of two major advantages over other traditional techniques. First, it captures moment-to-moment changes in eye gaze during natural reading and viewing without the need to rely on secondary tasks (e.g., providing verbal responses) (Conklin & Pellicer-Sánchez, 2016). Second, eye-tracking provides a rich record of learners' reading/viewing behavior (e.g., location, length, and sequence of eye movements), which allows for quantifying their visual attention allocated to a target linguistic form or a region (Conklin & Pellicer-Sánchez, 2016) and therefore provide valuable information about learners' cognitive operations when interacting with visual stimuli (Just & Carpenter, 1976). Eye movement data include two major types: fixations and saccades. Fixations refer to "periods during which the eye is relatively still, and the individual is looking at a specific area in the visual field" (Godfroid, 2020, p. 31). During most fixations, people extract and process information from a specific area they are looking at, which is referred to as the point of gaze or the point of regard. Fixation durations normally range from approximately 50 milliseconds to over 500 milliseconds (Rayner, 1998). Saccades refer to very fast movements of the eye between two eye fixations, which bring the eyes from one area to the next to identify new visual information (Godfroid, 2020).

Conklin and Pellicer-Sánchez (2016) defined and described eye-tracking measures that are frequently used in the field of L2 reading and how they are calculated. An example sentence with an area of interest (AOI, also referred to as region of interest, or ROI) shaded in grey is presented in Figure 9. Fixations are indicated by circles, and the number indicates their order. Fixations are often reported in terms of the number of fixations within an AOI, the amount of time spent in an AOI (i.e., fixation duration), and the probability of fixating an AOI (Conklin & Pellicer-Sánchez, 2016). Besides, eye-tracking data is often reported in terms of "early" and "late" measures (e.g., Altarriba et al., 1996; Staub & Rayner, 2007). Early measures are assumed to reflect viewers' initial stage of processing, such as word recognition in reading (Conklin & Pellicer-Sánchez, 2016; Godfroid, 2020). Commonly used early measures include first fixation duration, first pass reading time, and likelihood of skipping. Late measures, on the other hand, tap into viewers' strategic processes and may signal more effortful processing (Conklin & Pellicer-Sánchez, 2016; Godfroid, 2020). Second-pass reading time, rereading time, total reading time, and fixation count are frequently used as late measures. Depending on different research topics and aims, the classification of early and late measures might be slightly different (Godfroid, 2020). It is generally accepted that the use of several different measures may provide a fuller picture of learners' cognitive processes, although the measures are not completely independent from one another (Conklin et al., 2018; Godfroid, 2020; Rayner, 1998).

Figure 9 Illustration and Definitions of Frequently Reported Eye-Movement Measures in Text-Based Studies (Adapted from Conklin & Pellicer-Sánchez, 2016, p. 456)



Processing stage	Measure and depiction on figure	Definition and description
Early	Likelihood of skipping: 0 (the AOI is fixated)	– The likelihood that an AOI is skipped (not fixated at all) during the first time it is encountered – Some software packages indicate if an AOI was skipped (e.g., Data Viewer outputs 1 = AOI skipped and 0 = fixated) – Alternatively skipping can be calculated manually; in the AOI the number of trials with no first fixation duration (equals 0) is divided by the total number of trials
Early	First fixation duration: [3]	The duration of the first, and only the first, fixation on an AOI
Early	First-pass reading time (or gaze duration if the AOI is a single word): [3+4]	The sum of all fixations on an AOI before exiting to the right or left
Late	Second-pass reading time: [6+9]	Sum of fixations on an AOI after it has been exited for the first time
Late	Total reading time: [3+4+6+9]	Sum duration of all fixations on an AOI
Late	Fixation count: 4 fixations	Total number of fixations of an AOI

For text-based studies, eye-tracking serves as a means of gauging processing load (Tanenhaus & Trueswell, 2006) based on the eye-mind link assumption that longer eye fixation durations may indicate more processing or an increased level of task demands (Godfroid, 2020). In contrast, eye movements in image- or video-based studies are considered a representational measure, which relates to "when and where people fixate as the utterance unfolds" (Tanenhaus, 2007, p. 318). A greater degree of attention paid to an AOI may suggest more saliency and attraction, or heightened processing demands (Conklin et al., 2018). Similarly, there is not any standard as to which eye-movement measures should be used in image- or video-based studies. Importantly, the selection of eye-movement measures should always be driven by research aims and specific tasks of each research (Conklin et al., 2018; Godfroid, 2020). The measures frequently reported in existing video-based studies are summarized in Table 5. For a comprehensive review of eye-tracking in L2 research, see Conklin and Pellicer-Sánchez (2016), Conklin et al. (2018), and Godfroid (2020).

Table 5 Definitions of Frequently Reported Eye-Movement Measures in Video-Based Studies

Measure	Definition
Total fixation duration	Sum of all fixation duration on an AOI
Mean fixation duration	Average duration time of fixations on an AOI
Fixation count	Total number of fixations on an AOI
Run count (or visit count)	Total number of visits made to an AOI
Percentage of fixation duration	Proportion of total fixation duration spent on an AOI
Skip rate	Likelihood that an AOI is skipped (not fixated on at all)

2.6.3 Eye-tracking Studies on Multimodal Processing

In the field of L2 assessment, a growing number of L2 studies have begun to adopt eye-tracking technology to investigate test-takers' viewing behavior on video-based tests. Suvorov (2015) investigated how test-takers interacted with context-based and content-based videos from the Video-based Academic Listening Test (VALT) with 33 university students. The test consisted of three context videos (i.e., videos only showing the lecturers and the setting), three content videos (i.e., videos containing visual aids, such as graphs or images), and 30 multiple-choice questions. An eye tracker was used to capture test-takers' eye movements during test completion. Note-taking was allowed in the process. AOIs were identified as areas where the video was played during the VALT, and eye-tracking data were analyzed using three global measures: fixation rate (the number of fixations per second), dwell rate (the number of visits to an AOI per minute), and the percentage of total dwell time (the percentage of time that a viewer spends looking at an AOI). The results showed that test-takers attended to the visual information in the content videos at a higher rate than in the context videos, and there was no significant relationship between the eye-tracking measures and test scores.

Suvorov (2018) reported the results of qualitative data collected from the same 33 participants. Recordings of their eye movements while completing the VALT were used as stimuli, and they were asked questions regarding the aspects of visual input that they found helpful or distracting. The results showed that test-takers mainly focused on characteristics of the lecture instructor (e.g., appearance, gesture, and actions) and elements related to the lecture content (e.g., visual aids and textual information displayed) during viewing. Lecture-related visuals were more frequently mentioned as helpful than instructor-related visuals, as visual aids and textual information facilitated lecture comprehension and note-taking. Some participants, however, considered instructor-related aspects useful because seeing the instructor's mouth improved their comprehension, and the instructor's presence helped them maintain focus on the lecture.

On the other hand, test-takers reported that the instructor's body movements distracted them from listening and/or note-taking, and lecture-related elements, such as visual aids, were difficult to interpret or to be integrated with audio content.

While Suvorov's (2015, 2018) experiments explored how test-takers interacted with different types of videos, Batty's (2020) study focused specifically on test-takers' visual attention to speakers' social cues on a video-mediated listening test. Twelve Japanese undergraduate students of L2 English watched six short videos selected from a television program, featuring two actors having a face-to-face daily conversation. An eye tracker was used to record participants' eye movements during video viewing, and note-taking was allowed in the process. Test-takers then completed one multiple-choice item for each video, followed by a cued retrospective interview with their eye-gaze recordings as stimuli. They were asked to recall and verbalize reasons for their behaviors while completing the test. Due to the unavailability of advanced eye-tracking software and challenges of creating dynamic AOIs that could capture continual movements, scanpath-overlaid videos of test-takers' viewing behavior were manually coded based on oculomotor events of interest (e.g., scanning the speaker's or listener's face and looking at their gestures) at 0.10-second intervals. The result showed that test-takers spent most of their time (81.74%) looking at the speaker's face and largely split their attention between the speaker's eyes and mouth. Gestures seemed to attract little attention during video viewing, although being reported by Suvorov (2018) as a useful component of instructor-related visuals.

Research carried out by Suvorov (2015, 2018) and Batty (2020) demonstrated the value of triangulating eye-tracking with stimulated recall data, providing novel insights into how L2 test-takers use visual information during multimodal listening tests. Suvorov (2018), in particular, uncovered various instructor-related and lecture-related aspects of visual information that test-takers focused on and found helpful or distracting for lecture comprehension. Nonetheless, Suvorov's experiment did not directly compare the helpfulness of instructor-related and lecture-related visuals, and it is still unclear how learners integrate auditory and visual input to achieve comprehension. The only study combining aural and written texts with pictures to explore adult L2 learners' processing of multimodal input was conducted by Pellicer-Sánchez et al. (2021), with a specific focus on L2 reading.

In this study, 25 L2 advanced learners of English and 22 L1 English speakers either read or read while listening to a story while their eye movements were recorded. The story was accompanied by 31 pictures illustrating the characters and actions described in the story. Immediately after reading, the participants were asked to complete text-related and picture-related comprehension questions. Areas surrounding the text and the picture were identified as AOIs. Four eye-movement indices were analyzed, including the percentage of the sum of all fixation durations within each AOI, the percentage of the total number of fixations within each AOI, average fixation duration within each AOI, and the number of integrative saccades between text and images. The results revealed that L2 learners who read while listening to the story spent significantly more time looking at the pictures and made a higher proportion of integrative saccades. Both reading conditions, however, led to a similar level of comprehension, and no difference was found between the L1 and L2 readers' allocation of attention to text and pictures. An interesting finding is that processing time on the text was negatively related to L1 readers' comprehension but was positively associated with better comprehension for L2 readers. A positive relationship between the L1 readers' processing time on images and comprehension was also discovered.

2.6.4 Eye-tracking Studies on Vocabulary Acquisition

Whilst a limited number of studies have explored learners' processing of multimodal L2 input, a major line of research has centered on the processes underlying vocabulary acquisition from reading. The study carried out by Godfroid et al. (2013) was among the first eye-tracking studies to contribute to this strand of research. Twenty-eight EFL learners at an advanced level were instructed to read 20 short paragraphs containing either pseudowords, known control words, or both, while their eye movements were recorded by an eye tracker. Participants were then presented with the same paragraphs, but target pseudowords were removed. They had to fill the gap with one of the 18 options provided. Fixation duration on the target items was selected as an eye-tracking measure. The results showed that the participants recognized an average of 23 % of the target items, and the longer they fixated on the target words during reading, the more likely they recognized them in the post-test.

Pellicer-Sánchez's (2016) study adopted a more sensitive vocabulary knowledge measurement, that is, a dimensions approach, to reveal subtle increases in incidental vocabulary acquisition from reading. Twenty-three advanced L2 learners of English were instructed to read a short story containing eight repetitions of six pseudowords while their eye movements were captured by an eye tracker. Four eye-tracking measures were adopted, including first fixation duration, gaze duration, number of fixations, and total reading time. A form recognition, a meaning recall, and a meaning recognition test were used to measure participants' knowledge of the target items. Results showed that the participants recognized the form and meaning of 86% and 75% of the pseudowords, respectively, and recalled the meaning of 55% of the pseudowords. The study also revealed that the pseudowords were processed significantly faster after three to four exposures and were processed in a similar manner to previously known real words after eight exposures. In addition, accumulative reading times spent on the pseudowords significantly predicted scores in the meaning recall test. No relationships were found between accumulative reading times and form or meaning recognition.

In contrast to Godfroid et al.'s (2013) and Pellicer-Sánchez's (2016) studies which used experimentally manipulated materials, Mohamed (2017) utilized authentic texts as the reading stimuli to look into the cognitive effects of word repetition on vocabulary processing and acquisition. Forty-two advanced L2 learners of English were instructed to read a graded reader containing 20 pseudowords and 20 known words that were repeated several times (number of occurrences ranged from 1 to 30) while their visual attention to the target items was captured using an eye tracker. Three vocabulary post-tests (i.e., form recognition, meaning recall, and meaning recognition) were administered to assess the participants' knowledge of the pseudowords. A variety of eye-tracing measures were employed, including first fixation, gaze duration, total reading time, skip rates, regressions-in, and regressions-out of the AOIs. The result showed that the participants could retain the form of 42% of the pseudowords while recognizing and recalling the meaning of 30% and 13% of the pseudowords, respectively. Analyses of eye-movement data revealed that the participants spent more time on pseudowords than on known words, and their fixation durations decreased across exposures with more attention paid to the target words during early exposures. The amount of attention paid to pseudowords, as reflected in total reading times, was found to be positively related to learning outcomes in all three vocabulary tests.

Godfroid et al., (2018) also adopted authentic material to examine how learners' processing of unknown words changes with repeated encounters and whether repeated encounters predict vocabulary learning. Thirty-five advanced L2 learners of English read five chapters of an authentic novel containing 29 foreign words (i.e., *Dari* words) while their eye movements were recorded by an eye-tracking camera. After the treatment, participants completed three vocabulary post-tests (form recognition, meaning recall, and meaning recognition) measuring their knowledge of the foreign words. Only one eye-tracking measure, total reading time, was used in the study. Results showed that the participants could recognize the form and the meaning of approximately 30% of the foreign words, but they were only able to recall the meaning of about 13% of them. The analyses of eye-movement data showed a non-linear decrease in the processing times of foreign words: the processing times decreased rapidly from the first to the fourth exposures, followed by a slight increase, and a final gradual decrease. The study also revealed that longer processing time had a stronger relationship with the learning of word meaning.

One further eye-tracking study that employed authentic reading materials to explore vocabulary acquisition from reading is the work of Elgort et al. (2018). Additionally, Elgort and colleagues employed low- and high-frequency real words as target items to investigate whether the lexical representations generated from the reading activity could be accessed in other semantically-neutral contexts. Forty Dutch-speaking L2 learners of English at a higher intermediate to an advanced level were instructed to read an expository text. After the reading activity, the participants were asked to read a set of semantically-neutral sentences that contained the target words. An eye tracker was used to record the participants' eye movements during the reading of the text and the sentences. Six eye-tracking measures were adopted to gauge the processing time of target words, including first fixation duration, gaze duration, total reading time, go-past time, number of fixations, and number of regressions. A meaning recall test was then administered measuring the participants' ability to recall meanings of all target items. The results showed that the participants could recall the meaning of 34% and 99% of the low-frequency words and control words, respectively. The analyses of eye-tracking data indicated that learners spent a longer time processing the low-frequency target words than the high-frequency target words. In line with previous findings, differences in the processing time decreased significantly by the eighth exposure. In addition, there were differences in the processing of target items in the neutral sentences and in the last occurrence of the expository text. The participants' fixations and reading times were significantly longer in the former context than in the latter one.

Although these eye-tracking studies used different types of target items (i.e., pseudowords, foreign language words, and low-frequency real words) and reading materials (i.e., reading materials developed for the purpose of the studies and authentic reading materials) to investigate incidental vocabulary acquisition from reading, four consistent patterns have been reported: (a) learners' ability to acquire vocabulary knowledge from reading; (b) longer processing times on novel words than on familiar words, (c) an overall downward trend of processing time on novel words with repeated exposures, and (d) a potentially positive link between total reading time and vocabulary acquisition.

Another strand of eye-tracking studies has focused on incidental vocabulary acquisition from viewing captioned videos. Montero Perez et al. (2015) investigated L2 learners' processing of novel French words in two types of captions (full captioning and keyword captioning) and the role of text announcements (with or without announcements of an upcoming vocabulary test). Fifty-one Dutch-speaking undergraduate students watched two authentic videos with either full captioning or keyword captioning under either intentional or incidental learning conditions. An eye tracker recorded the participants' eye movements during viewing. Eighteen authentic words were chosen as target items, and participants' prior knowledge of the words was pre-tested. After the treatment, the participants' vocabulary gains were measured by four tests (i.e., a form recognition, a meaning recall, a meaning recognition, and a clip association). The clip association test required participants to indicate whether they could associate the words that they thought had appeared in the videos with the corresponding video, which aimed to control for guessing on the form recognition test. The comprehension test was administered after video watching only to ensure that the participants watched the videos attentively. Three eye-tracking measures were adopted to indicate the amount of attention directed to target items: gaze duration, second-pass reading time, and total fixation duration.

The results of eye-movement data showed that the intentional group spent significantly more time on the target words than the incidental group. When being forewarned of the vocabulary post-test, the keyword captioning group exhibited longer processing time on the target words than the full captioning group. In addition, longer second-pass reading time and total fixation duration were found to be positively related to the form recognition test scores for the full captioning, intentional group. Interestingly, for the full captioning, incidental group, a negative correlation between second-pass reading time and the form recognition test scores was found. This contradictory finding was interpreted by the researchers as suggesting that second-pass reading time might reflect processing problems rather than learners' increased intention to learn a word (e.g., guessing the meaning from context and encoding it in memory).

More recently, Montero Perez (2019) examined the effect of pre-learning vocabulary on learners' attention allocation during multimodal viewing. Thirty intermediate Dutch-speaking L2 learners of French watched a captioned video embedded with 10 pseudowords. Five pseudowords were pre-learned through word pairs (i.e., matching the L1 translation with the target words) while the other half only appeared in the video. After viewing the video, a form recognition and a meaning recognition test were used to measure the participants' knowledge of the pseudowords. Their eye movements during video watching and test completion were registered by an eye tracker. Four eye-tracking indices were used, including first fixation duration, total fixation duration, fixation count, and regression count. The results revealed that pre-learning did not affect the participants' eye fixations on the target item in terms of any of the eye-tracking measures. However, more skipping behaviors were observed for the pre-learned words than for the unknown pseudo-words. In addition, the results showed that pre-learning did not affect the time spent on the caption or the image area, but it had a significant impact on the participants' test completion behavior: they spent more time on test items for unknown pseudo-words than pre-learned pseudowords.

Apart from the investigation on the role of L2 caption in vocabulary acquisition from viewing, Wang and Pellicer-Sánchez (2022) further looked into learners' processing of unknown words in L1 video subtitles. In this experiment, 112 intermediate to advanced Chinese students of L2 English watched a short documentary in four types of subtitle conditions (L1, L2, bilingual, or no subtitles) while their eye-movement were recorded by an eye tracker. Vocabulary pre-tests and post-tests were administered to measure the participants' knowledge of 24 real words for form recognition, meaning recall, and meaning recognition. Three early eye-movement measures (first fixation duration, first-pass reading time, and skip rate), two late-processing measures (second-pass reading time and second fixation), and two aggregate late measures (total reading time and fixation count) were examined. The results revealed that although bilingual subtitles did not show any advantage for the learning of word forms, the participants using bilingual subtitles outperformed those using L2 subtitles in meaning recognition and outperformed the ones using L1 subtitles in meaning recall. It was also reported that participants in the bilingual subtitles condition spent significantly more time reading target words' L2 translations than the words themselves. Longer processing time on the target words significantly predicted gains in form recognition for the bilingual and L2 subtitles groups, meaning recall for the bilingual subtitle group, and meaning recognition for the L2 subtitles group.

Taken together, these eye-tracking studies on vocabulary acquisition from L2 viewing seem to yield inconsistent findings regarding the relationship between eye-tracking measures and vocabulary gains. Considering that eye-tracking data on their own might not be able to explain these distinctive effects, some L2 vocabulary researchers have triangulated data from eye-gaze recordings, vocabulary tests, and stimulated recall to explore how attention and awareness contributed to incidental vocabulary learning from reading. In Godfroid and Schmidtke's (2013) experiment, 29 advanced EFL learners read 20 English paragraphs embedded with 12 pseudowords while an eye tracker recorded their eye movements. Learners' eye fixation duration was used as a measure of attention. The participants then took an unannounced vocabulary post-test. The pseudowords were removed from the original sentences, and the participants had to fill the gap with 18 options provided. Next, the participants took part in an interview in which they were presented with the vocabulary post-test and were required to indicate whether they remembered what they had answered and whether they could remember reading the word. If they did, the researcher would ask what they were thinking about while reading. Results showed that the participants could recognize an average of 2.1 out of 9 words, and the pseudowords that the participants could remember having read in the paragraphs were fixated longer. In addition, both longer fixation time and recollection of reading the pseudoword predicted better word recognition.

2.6.5 Summary

This section has introduced two techniques, verbal reports and eye-tracking, for collecting information about processes underlying multimodal comprehension and vocabulary acquisition. The joint application of stimulated recall and eye-tracking has been shown to help researchers gain information not only about learners' visual attention but also about conscious thought processes during exposure to multimodal L2 input. In addition, these process measures combined with offline comprehension tests and vocabulary post-tests assessing different aspects of lexical knowledge allow researchers to paint a fuller picture of the relationships between processing and both comprehension and acquisition. Hence, a mixed-method design appears to be an appropriate approach to obtain a fine-tuned picture of learners' processes during L2 viewing and learning.

2.7 Study Aims and Research Questions

In view of the previous research discussed above, the current study intended to expand existing research on task repetition within the framework of TBLT. In particular, it aimed to investigate the benefits of repeated performance of a video-lecture-based task on learners' processing of multimodal L2 input and acquisition of technical vocabulary. An additional aim of the current study was to examine the impact of task repetition on the relationship between learners' multimodal input processing and both lecture comprehension and vocabulary acquisition. To achieve these goals, data was collected from multiple sources, including eye-gaze recordings, stimulated recall interviews, and offline measures (i.e., a free recall test and a set of vocabulary post-tests). It was in the hope that by capturing listener's real-time lecture viewing behavior through eye-tracking, gaining information about their conscious cognitive activities during viewing through stimulated recall, and measuring their learning outcomes by vocabulary post-tests, a more comprehensive understanding of L2 learners' attentional (conscious or unconscious) and acquisitional processes during multimodal lecture viewing could be achieved. To the best of my knowledge, this study constitutes one of the first attempts to triangulate these various sources to investigate L2 learners' multimodal processing and vocabulary acquisition from repeated task performance. To guide the present study, the following research questions (RQs) were formulated:

To what extent does repeating a multimodal lecture-viewing task affect:

1. learners' visual attention to the lecture instructor and labeled diagrams, as reflected in their eye movements?
2. learners' multimodal listening processes, as reflected in their stimulated recall comments?
3. the relationship between learners' visual attention to the lecture instructor as well as the labeled diagrams and their lecture comprehension?
4. learners' incidental acquisition of technical words, as measured by offline vocabulary tests?
5. learners' visual attention to technical words, as reflected in their eye movements?
6. learners' awareness of technical words, as reflected in their stimulated recall comments?
7. the relationship between learners' visual attention to technical words and their vocabulary acquisition?

Following previous TBLT research on the effects of task repetition on L2 oral production (e.g., Gass et al., 1999), task repetition was operationalized as performing the same multimodal lecture-viewing tasks three times. A set of hypotheses for each research question were formed. For learners' visual attention to lecture components (RQ1), it was hypothesized that learners' attention to the instructor would increase, whereas their attention to the diagrams would decline during repeated viewing. This was because diagrams might be more helpful to facilitate comprehension (Suvorov, 2018) in initial viewing, and learners' attention would naturally shift to the speaker (Batty, 2020; Gullberg & Holmqvist, 2006) after having gained some familiarity with the diagrams in subsequent viewing. Regarding learners' conscious cognitive processes (RQ2), based on the results of studies by Field (2015) and Holzknecht (2019), it was assumed that task repetition would affect processes involved in listening, reading, and visual comprehension, as well as the way learners integrate information presented in different modalities and the use of listening/viewing strategies. For the relationship between processing and lecture comprehension (RQ3), given the mixed findings yielded by existing studies (Pellicer-Sánchez et al., 2021; Suvorov, 2015), a nondirectional hypothesis was formed.

For vocabulary acquisition (RQ4), drawing on studies reporting a positive effect of task repetition on vocabulary learning (Ellis & Chang, 2018; Shintani, 2012a), it was assumed that participants who performed the task repeatedly would make greater vocabulary gains. In terms of online processing of vocabulary (RQ5), it was hypothesized that the time spent on the technical words would decline during repeated task performance, based on similar patterns found in L2 reading research that learners spent significantly less time processing unknown words after repeated exposure (Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016). In the absence of empirical studies on how repetition affects learners' awareness of target words (RQ6), following Skehan's (1998) limited capacity model, it was postulated that participants' level of awareness of target vocabulary would increase during task repetition. Lastly, considering the inconsistent findings yielded by previous studies for the relationship between visual attention and learning (Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022), the hypothesis for RQ4 is non-directional.

Chapter 3 Methodology

This chapter describes the methodology used in the present study. It begins with an explanation of the overall design, followed by a description of research ethics, participants, materials and instruments, as well as apparatus. The procedures related to data collection and analyses are then reported. The chapter ends with an overview of the statistical analyses.

3.1 Design

The current study adopted a mixed-methods research design, with 75 Chinese L2 users of English recruited through snowball sampling. Initially, I sent recruitment emails to postgraduate students enrolled in applied linguistics and TESOL programs, and the students who had participated in the study would refer other individuals from social science programs. The participants were assigned to a control group, (n = 30), a repetition group (n = 30), and a stimulated recall group (n = 15) using stratified random assignment based on a pre-administered listening proficiency test, namely, the Cambridge Certificate in Advanced English (CAE) test. This was to ensure that the three groups were as homogenous as possible in terms of their listening proficiency. Participants' receptive vocabulary size was measured by the 14k Vocabulary Size Test (VST, Nation & Beglar, 2007).

The control group performed a video-lecture-based task once, while the repetition group did the same task three times. The task asked participants to watch an introductory neurobiology lecture while taking notes. An instructor presented the lecture with the help of labeled diagrams representing biological structures. Eleven technical terms presented in the diagrams were selected as target words. The repetition group was not informed of the repeated viewing condition to ensure that the group could engage in the task naturally during the initial viewing, without altering their responses due to awareness of the repetition condition. By conducting a comparison between the repetition group's behavior during the first viewing and that of the control group, the homogeneity of the two groups was confirmed, which further strengthened the validity of the research design. However, both groups were made aware of an upcoming free recall test (i.e., retelling the lecture to a friend) before watching the lecture. Immediately after completing the treatment, the control and the repetition groups completed an unannounced vocabulary post-test, followed by the free recall test, a perception questionnaire, and a post-experiment questionnaire. The stimulated recall group was randomly assigned to three subgroups and performed the task once (n = 5), twice (n = 5), or three times (n = 5), respectively. They were then asked to describe their thought processes during their last task performance. An eye tracker was used to capture participants' visual attention to the instructor, diagrams, and target words during each viewing. Two weeks after the treatment, an unannounced delayed vocabulary post-test was administered to the control and the repetition groups to assess their retention of the target words.

3.2 Research Ethics

Given that the current research involved data collection from human subjects, research ethics approval was sought from the Research Ethics Committee at the Institute of Education (IOE), University College London (UCL). At the beginning of the study, each participant was given a general introduction and provided with an information sheet (see Appendix 1) explaining all aspects of the research, including (a) the main purpose of the study, (b) the overall research procedures, (c) the expected location and duration of participation, (d) the measures taken to maintain privacy and confidentiality, and (e) the participants' right to withdraw from the study at any time. Each participant's agreement to participate in the study was obtained by them signing a consent form (see Appendix 2). The original ethics application form, information sheet, and consent form were later revised to reflect new arrangements addressing the context of the COVID-19 pandemic. Updated elements included (a) moving all experimental sessions online except for eye-tracking sessions; (b) carrying out eye-tracking sessions following the UCL and government COVID-19 guidelines to guarantee safety for participants and researchers; (c) outlining potential risks of COVID-19 and additional safety measures in the ethics form, information sheet (see Appendix 3), and consent form(see Appendix 4); (d) arranging online meetings to answer questions and discuss any concerns participants might have prior to their participation; (e) requesting participants to confirm that they were aware of potential risks and understood the procedures in place for reducing COVID-19 transmission in the consent form. An IOE Fieldwork Risk Assessment form (see Appendix 5) and an IOE Postgraduate Research (PGR) Student Resuming Fieldwork proforma (see Appendix 6) were also submitted to provide further information concerning the hazards of COVID-19 in relation to (a) traveling to and from fieldwork; (b) fieldwork sites or setting, and (c) heightened risk for vulnerable groups.

3.3 Participants

In total, 134 Mandarin speakers of L2 English at a UK university (i.e., UCL) participated in the study. Forty-six students took part in the pilot study, whose aim was to check the suitability of materials, instruments, and research procedures. Lecture videos and the free recall test were piloted with 10 participants. They were asked to complete the task either once or three times while their eye movements were recorded by an eye tracker. This also helped the researcher practice using the eye-tracking system. Selected target words and vocabulary post-tests were tested with another 34 pilot participants, and a stimulated recall procedure was also piloted with two additional participants. For the main study, in total, 90 participants were recruited, but 15 were excluded. Four participants were excluded because their listening proficiency fell below the B2 band, the minimum level determined for the study in terms of the Common European Framework of Reference (CEFR). Two participants could not continue with the study because of difficulty with tracking their eyes, and data from three participants were discarded due to poor eye-movement data quality. Data from another six participants were removed from the analysis as they reported being familiar with the lecture content or the meaning of target items on the post-experiment questionnaire.

The 75 participants in the final sample for the main study were students enrolled in postgraduate social science programs. Eighty-four percent of participants were doing programs in applied linguistics, teaching English to speakers of other languages (TESOL), and education. There were 71 female and 4 male participants, with ages between 21 and 31 (M = 23.84, SD = 2.21, 95% CI [23.33, 24.35]). The three groups achieved comparable scores on both the CAE test and the VST (see Table 6). The large majority of their listening scores fell into the B2 and C1 bands, with a small percentage of participants (8%) reaching the C2 level, and all participants' receptive vocabulary size was above 6000 word families.

Table 6 Descriptive Statistics for the Participants' Proficiency Level

Group	CAE[a]		VST[b]	
	M (SD)	95% CI	*M (SD)*	95% CI
Control group (n =30)	19.10 (4.44)	[17.44, 20.76]	8827 (1300)	[8341, 9312]
Repetition group (n = 30)	19.03 (4.80)	[17.24, 20.82]	8850 (949)	[8496, 9204]
Stimulated recall group (n = 15)	18.60 (3.48)	[16.67, 20.53]	8860 (946)	[8356, 9384]

Note. [a]Cambridge Certificate in Advanced English test scores.
[b]Vocabulary Size Test scores.

3.4 Materials and Instruments

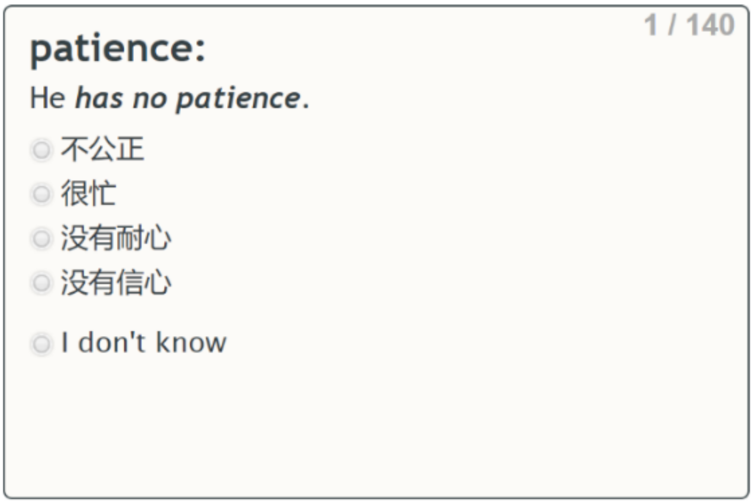
3.4.1 Listening Proficiency Test

I selected the listening section of a scored practice version of the CAE test to assess participants' listening proficiency because they were less likely to be familiar with this test than other popular proficiency tests, such as IELTS and Test of English as a Foreign Language (TOEFL). During the CAE test, participants were asked to fill in the blanks or to choose the correct answer from the choices given. They had the opportunity to hear each recording twice. The test was delivered in paper format, and it took about 40 minutes to complete.

3.4.2 Vocabulary Size Test

A bilingual version of VST (Nation & Beglar, 2007) provided by the Victoria University of Wellington was administered to participants after they completed the CAE test. The VST was selected for practical reasons: it is quick to administer and score. It was delivered online using a publicly available link (http://my.vocabularysize.com). The VST was comprised of 140 questions, with 10 items drawn from each of the first 14 frequency bands of 1,000 word families (See Figure 10). Vocabulary items were presented to participants both in isolation and in the context of a short sentence. Participants were asked to choose the correct meaning for a given item. As shown in Figure 10, all items were in English, but the choices were in the participants' L1. It took approximately 20 minutes to complete the test.

Figure 10 An Example of the Vocabulary Size Test Item



3.4.3 Multimodal Lecture-Viewing Task

The multimodal lecture-viewing task required participants to watch a videotaped lecture taken from an introductory course provided by the University of Chicago from the Coursera platform. While watching the lecture, participants' task was to take notes for a friend. They received the following written instructions prior to watching the lecture:

"Today, you receive a text message from your classmate Mary. She says she cannot attend a neurobiology lecture with you because she has an appointment with a doctor at the same time. She asks you to take notes for her so that you can explain to her what the lecture was about. The lecture will consist of three parts: (a) parts of neurons, (b) brain membranes, and (c) two types of photoreceptors of the human eye: rods and cones. After the lecture, you will leave a voice message for Mary. Using your notes, you will tell her as much as you can about the contents of the lecture. Now, you will watch the first part of the lecture. Please remember to take notes for Mary."

The lecture consisted of three videos featuring an L1 American English speaker presenting in front of a whiteboard with hand-drawn diagrams and keyword annotations. The first video introduced four basic parts of neurons, and participants could see a diagram of neuron structure (see Figure 11). The second video discussed the functions of brain membranes and the differences between the central nervous system (CNS) and the peripheral nervous system (PNS). Participants were presented with a diagram of three membranes that enclosed the brain (see Figure 12). Arrows in the illustration indicated the directions in which neural signals were conveyed. Two pictures of brain anatomy were also presented for a short amount of time. The last video introduced two types of photoreceptor cells in the human eye, that is, rods and cones (see Figure 13). The structure of photoreceptor cells and their preferred illumination were illustrated. There were four, four, and three words embedded in the respective videos. The target word did not repeat across the videos, except for two words that represented the main concepts in the first video but were mentioned once in the second video. While the target words were presented on the screen for most of the video duration, there were occasional instances where they were obscured because of the instructor's gestures, movements, or camera zooming in.

Figure 11 A Still Image of the First Video

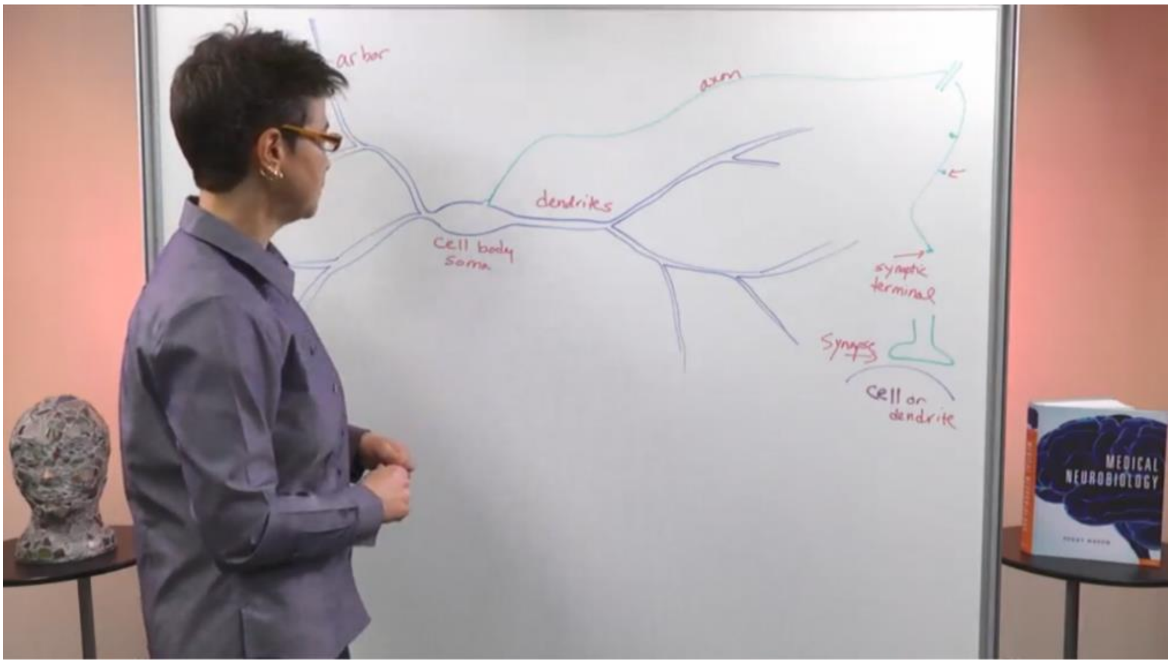


Figure 12 A Still Image of the Second Video

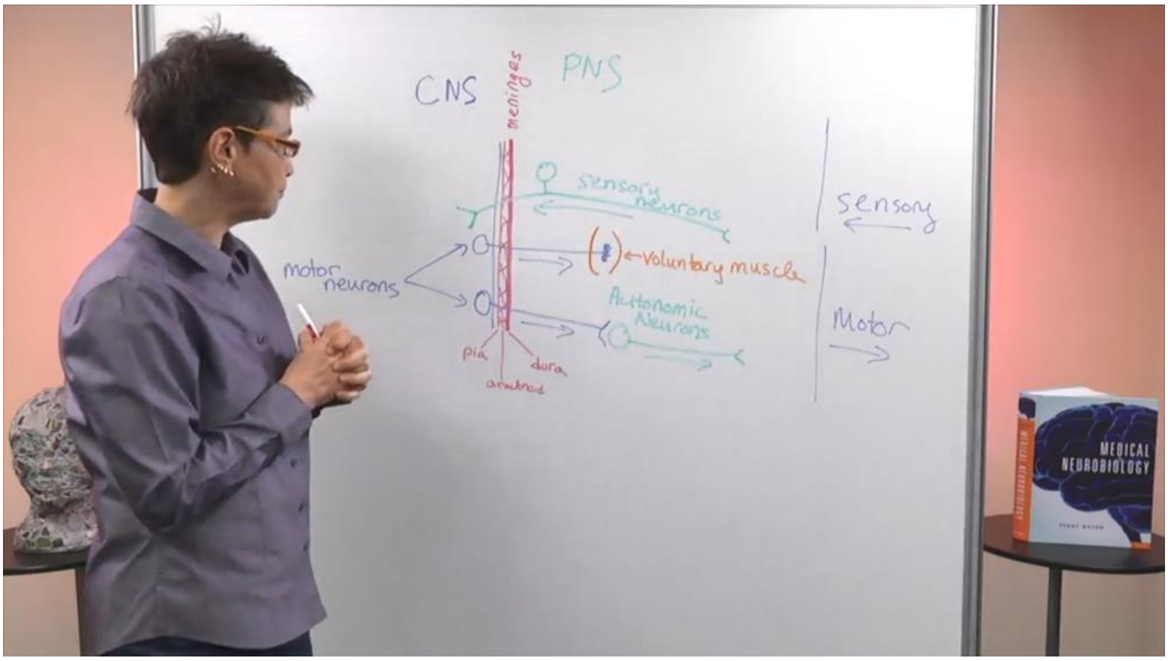
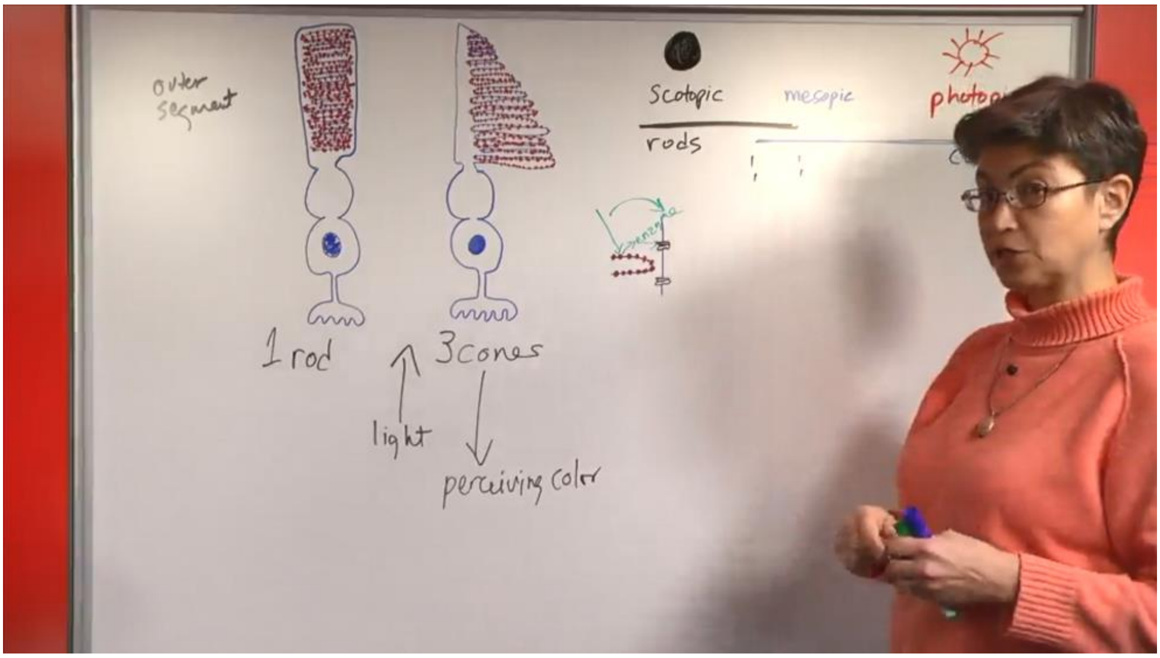


Figure 13 A Still Image of the Third Video



The rationale for selecting this lecture was threefold. First, the lecture was at a moderate level of difficulty, so it was assumed that participants would be able to comprehend the lecture content without much prior knowledge in the field. Specifically, the lecture covered some fundamental concepts of neurobiology using plain language, targeting a general audience. Apart from spoken texts that were easy to follow, it contained instructional diagrams representing biological structures and functions in a simple and clear manner. Diagram labels (1–3 words long) and keyword annotations also provided an appropriate amount of textual information. The results of the free recall test indicated that the lecture was comprehensible to both the control and the repetition groups. Second, the lecture was rich in content and was presented in multiple modalities, making it possible that the participants' interest could be sustained during repeated viewing. Lastly, considering the educational background of the participants, they were unlikely to be familiar with the lecture topics or the technical words presented. Nonetheless, the participants could potentially find value in learning knowledge about the brain and neurons, as neurobiology has many applications in social science, such as understanding the neural basis of behavior and cognition.

The three videos contained 30 frames per second (a frame lasted about 33 milliseconds). They were of approximately equal length (6 minutes), with each containing 770 running words on average. The Range program (Nation & Heatley, 2002) was used to assess the lexical profiles of the videos; the results revealed that knowledge of the first 3,000 most frequent word families provided a lexical coverage of 92 percent, beyond the 87 percent reported being adequate for successful viewing comprehension (Durbahn et al., 2020). Given the proficiency level of the participants, it was expected that they would have knowledge of the first 3,000 most frequent word families and therefore have no major comprehension difficulties. This was confirmed in the piloting of materials as well as the free recall test results. To retain task authenticity, the videos were not modified or simplified. They were only trimmed and combined to an appropriate length. Therefore, each video displayed characteristics of natural oral discourse, such as discourse markers (e.g., “You know”), false starts (e.g., “There's... the nucleus is here.”), and incomplete sentences (e.g., “And if we come over here… I have drawn out a model of one rod and one cone.”). For full video transcripts, please see Appendix 7.

3.4.4 Target Items

An initial bank of 12 possible target words was selected from the three videos based on the following criteria: they (a) were technical terms used in neurobiology and related to the main ideas presented; (b) were included in diagram labels so that learners would have access to their written forms; (c) were presented with hand-drawn illustrations depicting their concepts; (d) were accompanied by at least one oral explicit elaboration; and (e) presented concepts and forms unfamiliar to the participants both in their L1 and L2.

The steps of selecting target words for the present study are as follows. First, given that technical vocabulary forms an integral component of a subject-specific knowledge system, it could be identified using contextual information, using a technical dictionary, and referring to experts who have a good knowledge of the subject (Chung & Nation, 2004). Therefore, the first author started by identifying words that represented main ideas of the three videos. Specifically, target words *soma*, *dendrite*, *axon*, and *synapse* represented four basic parts of neurons and were selected for the first video. The second video introduced the brain membranes, specifically known as *meninges*, which comprised *dura*, *pia*, and *arachnoid*. Consequently, these four words were selected. The third video presented how two types of photoreceptor cells function under different lighting conditions, namely, *scotopic*, *mesopic*, and *photopic* conditions, so these words were chosen. Next, according to Chung and Nation (2004), the presence of definitions and labeling in diagrams are two main types of contextual cues that signal technical vocabulary. A further examination of the identified words ascertained that each word was accompanied by at least one aural definition and presented in written form as diagram labels to clarify specific components of illustrations (i.e., neural structures, the composition of brain membranes, and lighting conditions). I then checked all selected items in a neuroscience glossy (Purves et al., 2018) and consulted a doctoral researcher in neurobiology. It was confirmed that all target words were indeed listed in the glossary and related to the field.

Furthermore, only in cases where participants exhibited no prior knowledge of the target words in both their L1 and L2, the learning conditions involved the simultaneous acquisition of both new L2 forms and new subject knowledge. To ensure this, participants were recruited from social science programs and were asked about their enrollment in any courses related to neuroscience, neurobiology, cognitive psychology, and biological psychology. Only participants who reported having little knowledge of these fields were invited to the study. A perception questionnaire and a post-experiment questionnaire were administered to the participants at the end of the experiment, aiming to elicit indications of familiarity with the lecture topic or the meaning of any target words. Data from participants who reported being familiar with the lecture content or the target items were removed from the analysis. It was also observed that none of the stimulated recall participants used L1 translation of the target words while describing their thought processes, indicating the absence of any pre-existing familiarity with the words in their L1.

Using the same vocabulary tests as in the main study (i.e., form recognition, meaning recall, and meaning recognition), these items were then pilot-tested by 34 participants (i.e., students with similar backgrounds and from the same university as those who would participate in the main study). Based on the results of the tests, 11 items, known by less than three percent of the pilot participants, were chosen as target words for the main study. One item was removed because it was recognized by a few pilot participants.

To preserve the authenticity of the lecture, it was not possible to control for other aspects of the learnability of the target words, including part of speech (POS), word length, and frequency of occurrence. Item-level differences, however, were included as covariates in the statistical analyses. A detailed list of the target words is presented in Table 7. Among the 11 items, three were adjectives, and eight were nouns. The mean length of words was 6.3 letters, ranging from three to eight letters. Each target word appeared one to six times in the spoken texts. They were all low-frequency words (ranked above the 11,000-word level according to the British National Corpus (BNC). All word elaborations used in the videos could be classified as explicit, corresponding to what Flowerdew (1992) referred to as formal (i.e., precise statement of a word) and semi-formal (i.e., identifying key characteristics of a word) definitions. An example of a formal definition extracted from the video transcripts is: "The first is the cell body, also called the *soma*, and this is the part that all cells have. This is cell central." An example of a semi-formal definition is: "There is one *axon*. And while these dendrites are very local, they are gonna have a local distribution. This *axon* can go far, far distances." The instructor's gestures of pointing to a target item were considered as nonverbal signals. The relatedness of the 11 items to the field was also confirmed by an expert in neurobiology.

Table 7 Characteristics of the Target Items (in Alphabetical Order)

Item	Video prompt number	Length	POS[a]	FoO[b] in the spoken text	Number of elaborations	Number of nonverbal signals
arachnoid	2	9	n.	4	2	0
axon	1	4	n.	5	7	0
dendrite	1	8	n.	6	11	3
dura	2	4	n.	4	5	1
meninges	2	8	n.	3	6	2
mesopic	3	7	adj.	3	2	4
photopic	3	8	adj.	4	2	5
pia	2	3	n.	3	4	2
scotopic	3	8	adj.	3	3	11
soma	1	4	n.	1	10	1
synapse	1	7	n.	1	3	1

Note. n. = noun; adj. = adjective; [a]Part of Speech. [b]Frequency of occurrence.

3.4.5 Free Recall Test

The free recall test was employed to measure participants' lecture comprehension. The task was chosen for two main reasons. First, it was a communication-oriented task resembling what learners might do in real-life situations, that is, viewing lectures and recalling (Winke & Gass, 2016). Second, by analyzing recall protocols in comparison to original video transcripts, free recall tests could reveal misinterpretations, distortions, and inferences, making them an informative tool for measuring comprehension.

After viewing the lecture, participants had to leave a voice message for a friend, giving as much detail as possible about the lecture content. They were prompted by the notes they had taken. The following instructions were presented to the participants:

"This is the end of the lecture. Please leave a voice message to Mary telling her everything you understood and recalled from the lecture. You can use Mandarin, English, or a mixture of both languages. You can also use your notes to help you recall information whenever you feel necessary. There will be no time limit."

Given that the task aimed to assess comprehension, participants were allowed to use Mandarin, English, or a mixture of both. No time limit was set. The participants' recall protocols were captured using a digital recorder. It took approximately 8–15 minutes to complete the task.

3.4.6 Vocabulary Post-Tests

To obtain a multi-faceted picture of incidental vocabulary development, a set of untimed vocabulary tests (a form recognition, a meaning recall, and a meaning recognition test) was used to measure participants' knowledge of the target words. This set of tests was also used in the pilot study to identify the target words and served as immediate and delayed post-tests in the main study. As participants were unlikely to know the meaning of the target words, no pre-test was included in order to avoid testing effects. Participants' prior knowledge of the words, however, was checked in the post-experiment questionnaire. The immediate post-test was unannounced and administered right after the participants had completed their (last) task. Similarly, participants were not told in advance that they would have to take the delayed post-test measuring their longer-term retention of target items. The time interval between the immediate and delayed post-tests was set to two weeks in order to diminish any effects of reactivity. The form recognition and meaning recognition tests were constructed using the E-prime 2.0 software (Schneider et al., 2002). The meaning recall test was delivered in paper format.

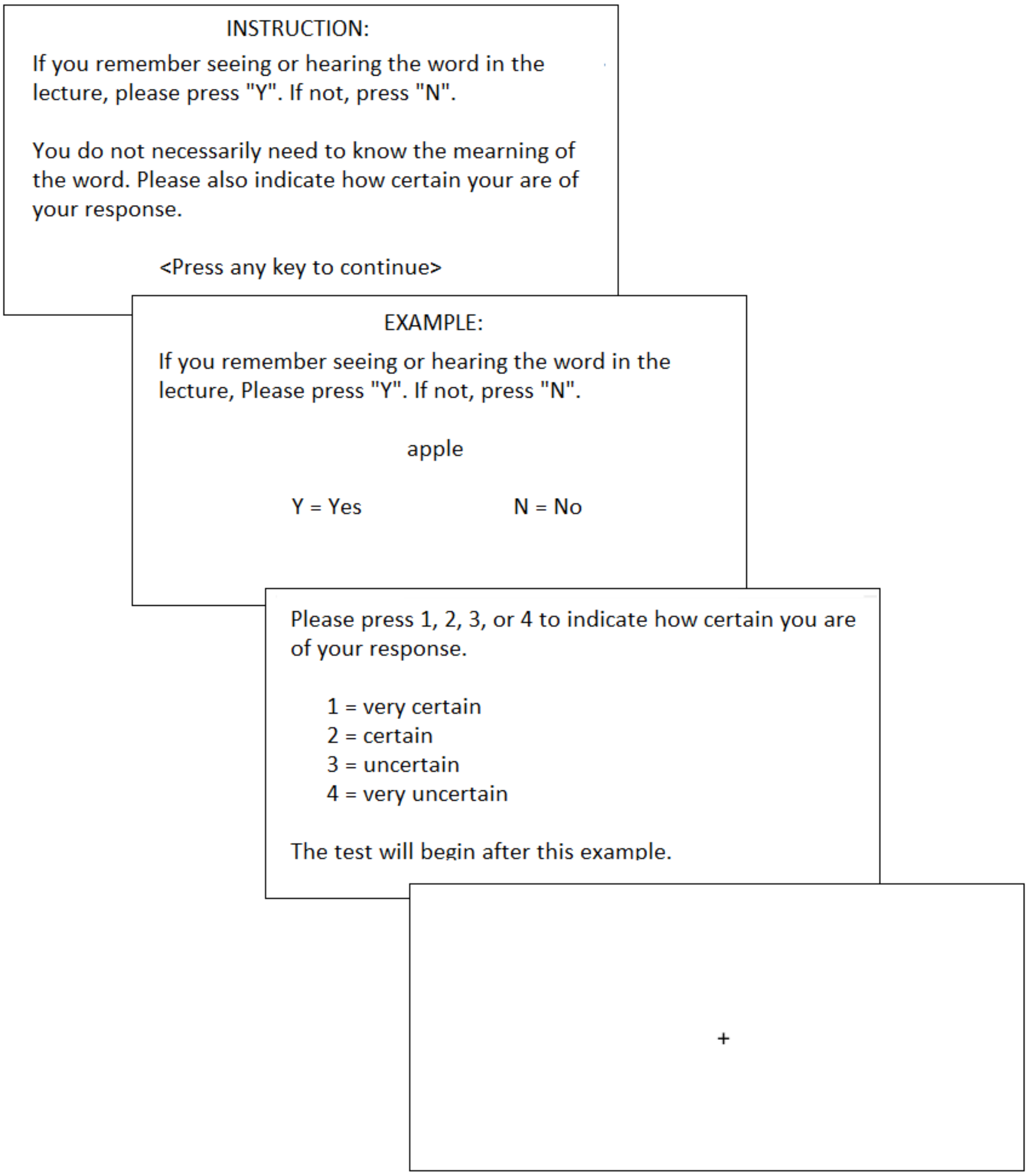
In the form recognition test, 11 target words together with 11 distractors were randomized and presented one by one to participants in both written and spoken forms. Recordings of the target words and distractors were prepared by a female L1 American English speaker. Careful attention was paid to the selection of distractors. They were neurobiology and medical terms, which were found to be unfamiliar to 99 percent of the pilot participants (see Table 8). The distractors also contained 2–3 syllables and were of the same word class as the target words (three adjectives and eight nouns). Participants were asked to press either “Y (yes)” or “N (no)” key to indicate whether they remembered seeing/hearing the word in the lecture (see Figure 14). Each target word remained on the screen until the participant chose an answer. Then, a confidence rating task followed, asking participants to indicate on a Likert scale from 1 to 4 how certain they were of their response (1 = very certain; 2 = certain; 3 = uncertain; 4 = very uncertain). The confidence rating task was used to provide insights into learners' metacognitive processes when completing the vocabulary post-tests, but the results of the task were not presented in the thesis as they were beyond the scope of the current research. A fixation cross was constructed to appear on the screen for 500 milliseconds between each set (i.e., a form recognition item and a confidence rating task) in order to signal the next item. The form recognition test took approximately 5–7 minutes to complete.

Table 8 Distractors Used in the Form Recognition Test (in Alphabetical Order)

Distractors	Part of speech
allele	n.
amyloid	n.
enteric	adj.
glia	n.
gyrus	n.
ischemic	adj.
myelin	n.
prion	n.
syncope	n.
telomere	n.
thalamic	adj.

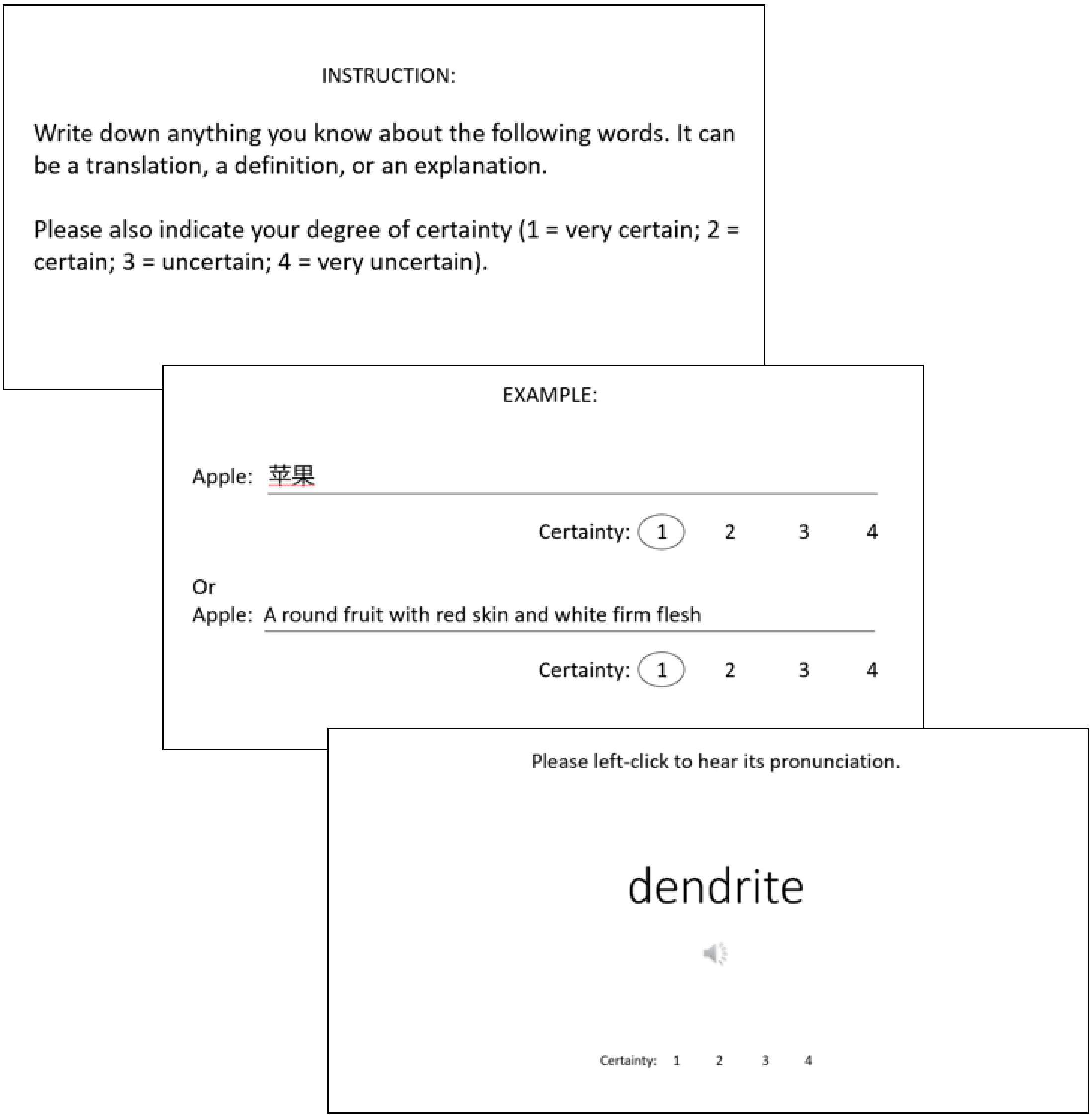
Note. n. = noun; adj. = adjective.

Figure 14 An Example of the Form Recognition Test



After completing the form recognition test, participants' knowledge of the meaning of the target items was measured with an untimed meaning recall test. Eleven target words were randomized and presented one by one to participants on a slide in their written and spoken forms (see Figure 15). Participants were asked to write down everything they knew about the meaning of the words presented on the slide. It could be a translation, an explanation, or anything else that demonstrated their knowledge in either their L1 or L2. Each participant had to click a play button to hear the spoken forms of words before writing anything down. Participants were also asked to indicate their degree of certainty as in the form recognition test. The test took about 5–15 minutes to complete.

Figure 15 An Example of the Meaning Recall Test

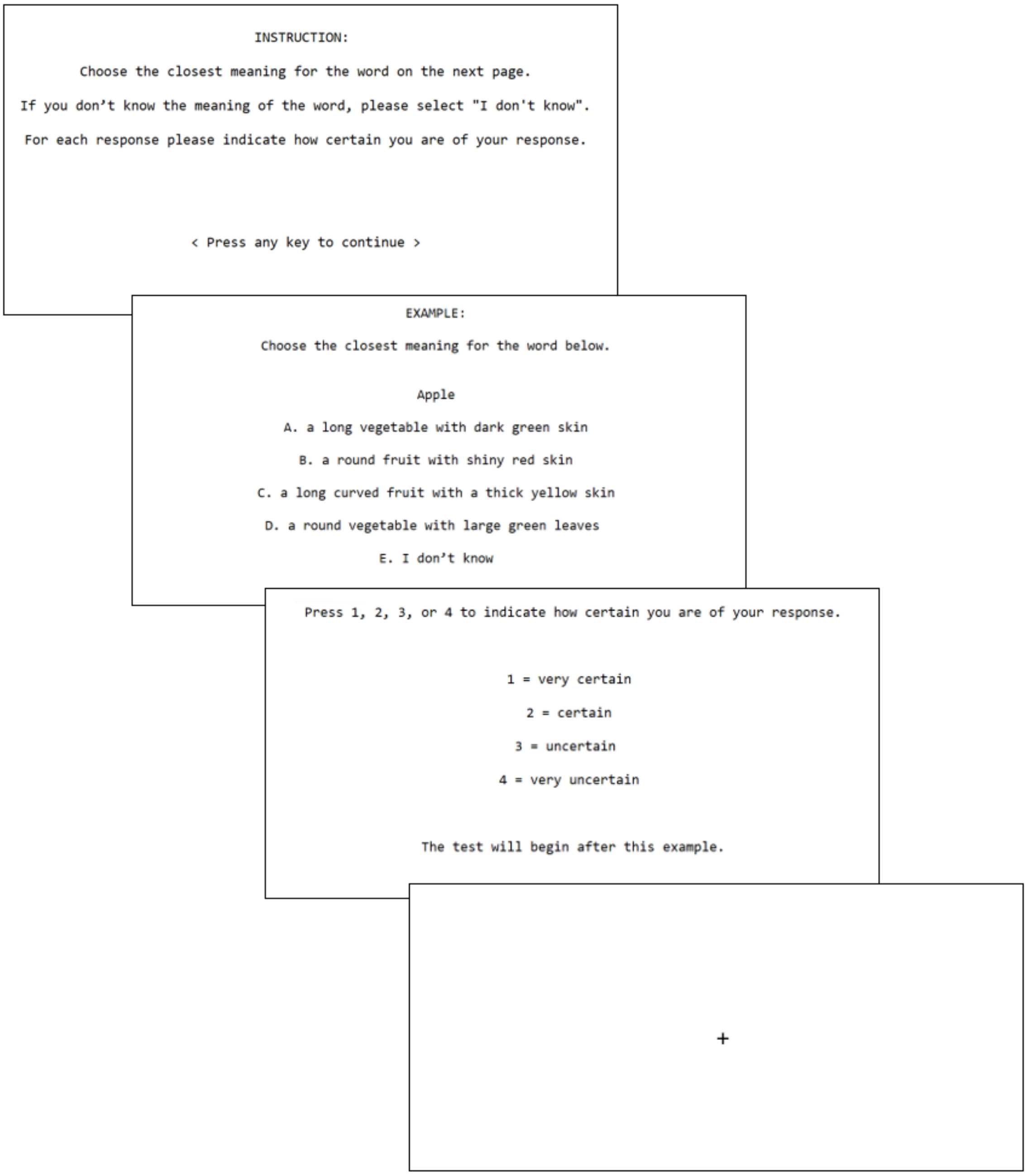


The meaning recognition test was administered last to capture knowledge below the level of meaning recall. It included multiple-choice items developed for each target word, consisting of five possible options: the correct meaning, a definition of a target word that appeared in the same video, a definition of a target word from a different video, a definition of a distractor, and an “I don’t know” option. To minimize guessing, participants were instructed to choose the “I don’t know” option when they did not know the answer. The 11 distractors were all semantically related to the content of the lecture (see Table 9). As in the form recognition test, the target words were randomized and presented in visual and auditory forms simultaneously, followed by a confidence rating task (see Figure 16). Participants had to choose the closest meaning for each target word by pressing the corresponding key. This test took about 5–7 minutes to complete. For the complete vocabulary post-tests, see Appendix 8.

Table 9 Distractors Used in the Meaning Recognition Test (in Alphabetical Order)

Distractors	Meaning
brain tumor	A mass of abnormal cells found in the brain
cerebrospinal fluid	The clear, colorless liquid found surrounding the brain
cognition	The process by which knowledge and understanding is developed in the mind
cornea	The transparent layer which covers and protects the outer part of the eye
cortex	The outer layer of the brain
fixation	The maintaining of the visual gaze on a single location
glia	The supporting cells of the central nervous system
perception	The way you notice things
saccade	Quick and simultaneous movements of both eyes
skull	The bone structure that forms the head and protects the brain
stroke	A sudden serious illness when a blood vessel in the brain bursts or is blocked

Figure 16 An Example of the Meaning Recognition Test



3.4.7 Questionnaires

Three questionnaires were administered to the participants, including a background questionnaire, a perception questionnaire, and a post-experiment questionnaire. All questionnaires were delivered in paper format. The background questionnaire was used to collect participants’ information about demographics and

English language learning experience, such as their age, gender, major, IELTS band score, and length of residence in English-speaking countries.

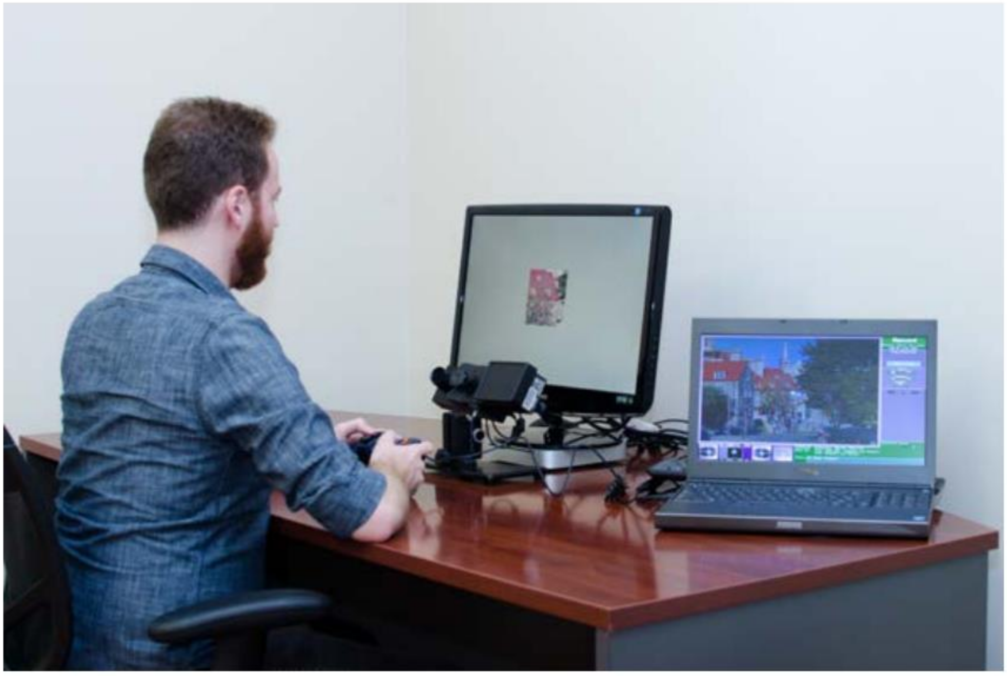
The perception questionnaire was administered to the participants immediately after they had completed the free recall test. It was purposely kept short and was delivered in simple English. The questionnaire included seven statements that participants needed to judge on a 5-point Likert scale, ranging from strongly disagree to strongly disagree. It was designed to assess participants' perceptions of (a) topic familiarity, (b) overall task difficulty, (c) the linguistic complexity of the lecture, (d) the cognitive complexity of the lecture, and (e) their ability to perform the task. The questionnaire responses were used to screen for participants who were familiar with the lecture topics.

After the perception questionnaire, a post-experiment questionnaire adapted from Winke et al. (2010) was administered to the participants measuring their prior knowledge of the target words. Participants had to indicate their familiarity with the target words on an unfamiliar-familiar continuum. This questionnaire was used to exclude participants who had prior knowledge of the target words. See Appendix 9 for the questionnaires.

3.5 Apparatus

The SR Research EyeLink 1000 Plus (2016) system, consisting of a desk-mounted eye tracker, a host laptop, and a display computer, was used to collect eye-movement data (see Figure 17). The eye tracker used infrared light to illuminate participants' eyes and recorded their pupil and corneal reflection to track eye movements. It was set in remote mode, as the participants had to take notes during the lecture-viewing task. In the pilot study, I found that the pilot participants' calibration results were always poor, although attempts were made to repeat the calibration routine several times. The poor calibration quality resulted from the default setting on the host laptop, which did not apply to the remote mode used in the study. After customizing the default screen setting (i.e., screen dimensions, display resolution, eye-to-screen distance, and camera-to-screen distance) on the host laptop, good calibration results were achieved. Moreover, although the system was by default configured to use a 16 mm remote lens, the pilot study found that a 25 mm remote lens provided better recording data, and the pupil was more efficiently detected with the 25 mm remote lens. Therefore, a 25 mm remote lens was employed in the main study, sampling the participants' right eye at 1000 Hz (i.e., every 4 milliseconds the system determined the participants' pupil position and size of the registered eye in relation to the computer screen) to provide more precise data. Eye-tracking data collected from the 10 pilot participants using the 16 mm remote lens were discarded.

Figure 17 The EyeLink 1000 Plus System (SR Research, 2016, p. 2)



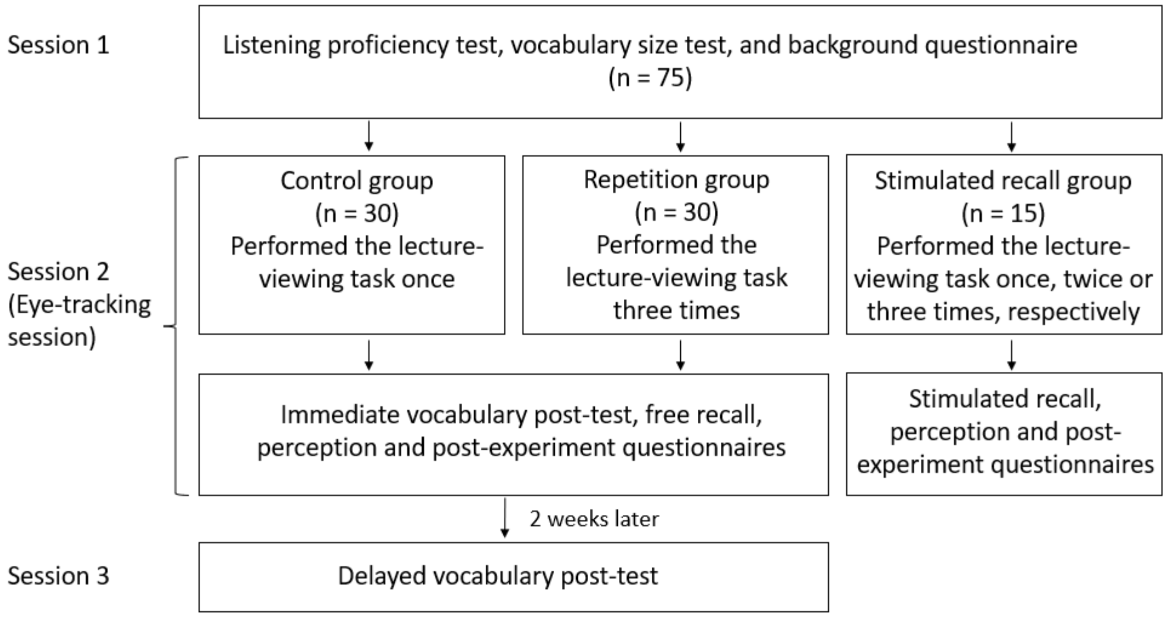
3.6 Data Collection Procedures

As shown in Figure 18, the data was collected over four weeks. In the first session, participants were provided with the information sheet and the consent form. The full intent of the study was not explained to the participants. Instead, they were told that the purpose of the study was to investigate multimodal lecture comprehension. They were then administered the CAE listening test, VST, and background questionnaire. This session was carried out face-to-face initially but was later moved online to Zoom because of COVID-19 restrictions. During the online session, participants were instructed to turn on their video cameras and share their screens to ensure that the research procedure was followed. The whole session took approximately 80 minutes.

The second, eye-tracking session took place at a language laboratory. Prior to the session, participants were informed of potential COVID-19 risks associated with the eye-tracking session and were encouraged to consider their participation carefully. Following the UCL and government guidelines, additional procedures and adjustments were put in place to minimize risks related to COVID-19 transmission during participants' visits to the laboratory. For example, participants and the researcher had to have a symptom-free COVID-19 lateral flow test (LFT) up to 24 hours before face-to-face sessions. When participants arrived at the entrance of the laboratory, temperatures were taken using a non-contact electronic forehead infrared thermometer, followed by a pre-experiment questionnaire gathering data on COVID-related symptoms (Appendix 10). The participants were required to sanitize their hands before and after the eye-tracking session and wear face masks during their whole visit. Surfaces and items used as part of the study (e.g., keyboard, mouse, headphone, etc.) were thoroughly sanitized after being used. Participants were instructed to watch the videos and take notes, while their eye movements were recorded by an eye-tracker. Prior to lecture viewing, they were informed that there would be a free recall test at the end. Participants who performed the task repeatedly were offered a five-minute break between repetitions to reduce fatigue. The total duration of the second session was between 70 to 120 minutes.

The third session took place two weeks after the treatment when the delayed vocabulary post-test was administered. Similarly, this session was carried out via a Zoom meeting since COVID-19 restrictions were implemented. Participants were also monitored through Zoom to ensure that they would not divert from the research procedure. The duration of the third session was approximately 20 minutes. At the end of the session, participants were informed of the full intent of the study.

Figure 18 Visual Diagram of the Research Procedure



3.6.1 Eye-tracking Procedures

The eye-tracking experiment was first constructed using the SR Research Experiment Builder (2011) software version 2.2.1. Eye-tracking data was collected using the EyeLink 1000 Plus (2016) system from one participant at a time in a quiet room. In preparation for the experiment, each participant was informed of the function of the eye tracker and an overview of the procedure. The participants then donned a headset and received several pieces of paper and pens for note-taking. At the beginning of the session, the participants were approximately 60 centimeters away from a 19-inch monitor with a resolution of 1920 × 1080. Once their seating position was adjusted, the participants were asked to complete a 9-point grid calibration. After watching a one-minute practice video, their eyes were calibrated again. Then, the instructions for the lecture-reviewing task were presented, and participants had to press the “Enter” key on the keyboard to start watching the first video and press the same key to proceed to the next video. Drift correction was performed before each video, and additional calibrations were carried out when necessary. I monitored participants’ eye movements during task performance in the inspection window on the host laptop. If a participant shifted outside the acceptable boundaries, their position would be moved slightly with minimum disruption.

3.6.2 Stimulated Recall Procedures

The stimulated recall procedures included two stages. First, the three subgroups of stimulated recall participants were shown a short example video of eye movements in order to familiarize them with the recall prompt. It was explained to the participants in everyday language that the pink circles in the video indicated their eye fixations. Next, the participants watched a replay of their eye-movement recording during their (last) viewing (see Figure 19). They were instructed to stop the recording at any time they wanted to verbalize what they were thinking while engaging in the task. If they did not do so, I stopped the video every 30 seconds and prompted the participants to describe their thoughts during task performance. The prompt questions were informed by previous studies on L2 listeners’ cognitive processes in listening assessment (Harding, 2011; Holzknecht, 2019; Holzknecht et al., 2017). Only general questions were asked, such as “What were you thinking when you were watching this part?” No responses were provided to the participants other than backchannelling cues. Participants were asked to clarify their comments only in cases where their responses were ambiguous or unclear, in order to achieve an accurate understanding of their intended meaning. This can be illustrated by the following quotation:

Participant: While listening to this part, she was talking about a "communication center". Then, I checked my spelling.

Researcher: Spelling of which word?

Participant: This "synaptic". Yes, and this "synapse" as well.

If participants responded that they could not remember, they were not encouraged or led to recall. All stimulated recalls were conducted in Mandarin, but the participants could also use English. The stimulated recall sessions were video-recorded using a camera to capture participants' spatial movements (e.g., pointing at the screen), which might enhance the interpretation and understanding of the stimulated recall data. The recalls were also recorded via a voice recorder in case the camera ran out of battery. The length of the individual stimulated recall interview varied from 36 to 58 minutes, with an average length being 44.5 minutes. The total duration of recall protocols gathered from 10 stimulated recall participants was 667 minutes. For full stimulated recall interview instructions, please see Appendix 11.

Figure 19 A Still Image of the Eye-Movement Recordings

3.7 Data Analyses

3.7.1 CAE Test

The CAE test contained 30 questions, and each question carried one mark. The official scoring criteria are presented in Table 10 (retrieved from https://www.cambridgeenglish.org/images/210434-converting-practice-test-scores-to-cambridge-english-scale-scores.pdf). The internal consistency of the test was assessed via Cronbach's alpha, resulting in an α of .72, which met the threshold for acceptability.

Table 10 CAE Listening Test Scoring Criteria

CAE listening practice test score	CEFR level
26	C2
18	C1
13	B2
11	—

3.7.2 Vocabulary Size Test

Participants' VST score was estimated by calculating the number of correct responses for all 140 items and multiplying the result by 100. The maximum score that could be achieved on the test was 14,000, indicating that a learner had a written receptive vocabulary in English of 14,000 word families.

3.7.3 Free Recall

The analysis of free recall protocols included two steps. Firstly, the transcripts of the three videos were subjected to idea-unit analysis, following Carrell's (1985) operational definition of an idea unit:

> Basically, each idea unit consisted of a single clause (main or subordinate, including adverbial and relative clauses). Each infinitival construction, gerundive, nominalized verb phrase, and conjunct was also identified as a separate idea unit.

In addition, optional and/or heavy prepositional phrases were also designated as separate idea units (p. 737).

For example, the following text extracted from the transcripts was separated into 14 idea units, with one of these units identified as repetitive.

// [Idea unit 1] “There are four parts to neurons. // [Repeated idea unit] Neurons have four parts. // [Idea unit 2] The first is the cell body, // [Idea unit 3] also called the Soma. // And this is the part // [Idea unit 4] that all cells have, this is cell central. // [Idea unit 5] There's… the nucleus is here // [Idea unit 6] with the DNA. // [Idea unit 7] This is like city hall, // [Idea unit 8] it's gonna give out all the orders. // [Idea unit 9] There's a manufacturing plant here // [Idea unit 10] where proteins are made. // [Idea unit 11] There is a power plant here. // [Idea unit 12] So this is really the place // [Idea unit 13] that keeps the cell going.” //

Based on these criteria, the three video transcripts were broken down into 437 idea units, with 143, 147, and 147 idea units in each video transcript. Given that the current study used authentic materials, two major types of units were excluded from the calculation: units containing disfluency features (e.g., repetition, false starts, and self-corrections) and units that were not directly relevant to the lecture (e.g., lecture introduction and previews of what was coming in the next video). Therefore, the resulting total was 338 idea units (75, 131, and 132 units for each video), which determined the highest possible score for the free recall test. Next, participants’ recalls were transcribed, checked, and coded for idea units using Carrell’s definition. Units including disfluencies and irrelevant content were again excluded. All participants recalled in English, except two who used Mandarin. Their non-English responses were translated into English. The recall protocols were scored regardless of spelling, grammatical mistakes, or language. Each correctly recalled idea unit was assigned 1 point. Incorrect idea units (misinterpretation and distortion) and idea units that were not mentioned in the transcripts (e.g., inference) were counted separately. To check interrater reliability, 20 percent of data were scored by a second coder (a postgraduate student undertaking doctoral studies in applied linguistics), yielding a high inter-coder reliability (Cohen's kappa = .89).

3.7.4 Vocabulary Post-Tests

The form and meaning recognition tests were scored in a binary fashion, with a correct answer getting 1 point and an incorrect answer getting 0. The maximum score for the form recognition test was 22 (including 11 distractors) and the maximum score for the meaning recognition test was 11. Drawing on Gablasova (2014), the meaning recall test was scored by calculating the proportion of core meaning components of the target words that could be successfully recalled by the participants. Only components that were essential for defining and distinguishing the target words from other keywords in the transcripts were considered core meaning components. For each target item, three to four core components were identified (see Table 11) based on video transcripts, definitions from a glossary of neurobiology terms, and the Cambridge Dictionary. If a participant recalled two out of three essential components, for example, the score would be 2/3. The meaning recall test was scored regardless of grammatical mistakes or language. Twenty percent of the data were scored by a second rater, a doctoral student in applied linguistics, resulting in high inter-coder reliability for the immediate post-test (Cohen's kappa = .84) and delayed post-test (Cohen's kappa = .82).

Table 11 Coding Scheme for the Meaning Recall Test (in Alphabetical Order)

Target item	Core component
arachnoid	1. Arachnoid is a layer of brain membrane/meninges.
	2. Arachnoid is spidery/looks like a spiderweb.
	3. Arachnoid is located between pia and dura/the middle layer of meninges.

Target item	Core component
axon	1. A neuron has one axon.
	2. Axon is responsible for sending out information.
	3. Axon can go far distances.
	4. Axon carries information to a terminal/communication center/synapse.
dendrite	1. Dendrites can branch many times/Neurons form dendritic trees.
	2. Dendrites are responsible for gathering information.
	3. Dendrites are extensions of cell body.
	4. Dendrites are very local/have a local distribution.
dura	1. Dura is a membrane/layer of meninges.
	2. Dura is tough/strong/hard.
	3. Dura protects our brains from injury/concussion.
	4. Dura is the farthest away from the CNS/the brain and spinal cord.
meninges	1. Meninges are the fence between the CNS and PNS.
	2. Meninges are made up of three membranes.
	3. The three membranes are pia, dura, and arachnoid.
mesopic	1. Mesopic is between scotopic and photopic/describes conditions where there is not too much nor too little light.
	2. Under mesopic conditions, we perceive muted colors.
	3. Both rods and cones are active in mesopic conditions.
photopic	1. Photopic describes bright conditions or conditions where we can see sunlight.
	2. Under photopic conditions, we perceive bright/vibrant colors.
	3. Cones are useful in photopic conditions.
pia	1. Pia is a membrane/layer of meninges.
	2. Pia is weak/thin/tender.
	3. Pia is the closest to the CNS/the brain and spinal cord.
scotopic	1. Scotopic describes dark conditions or conditions of little/dim light.
	2. Under scotopic conditions, we cannot perceive colors.
	3. Rods are useful in scotopic conditions.
soma	1. Soma refers to the cell body/cell central/the central part of a neuron.
	2. Soma contains nuclear with the DNA.
	3. Soma contains manufacturing plants and power plants.

Target item	Core component
	4. Soma keeps the cell going/maintains important cell functions.
synapse	1. Synapse is separated by a space/consists of a gap between two cells.
	2. Synapse is the point of information transfer/takes information and gives it to the next cell.
	3. The second cell can be various cell types (e.g., another neuron, muscles, and glands).

3.7.5 Stimulated Recall

The analysis of stimulated recall data started with transcribing the data in the language used by the participants (i.e., Mandarin with occasional English). The transcripts were also checked by a second transcriber, a Mandarin speaker of L2 English. The coding was then carried out on the Mandarin transcriptions to avoid data loss due to translation. To answer RQ2 (To what extent does repeating a multimodal lecture-viewing task affect learners' multimodal listening processes, as reflected in their stimulated recall comments?), the integrated model of multimodal comprehension described in Section 2.3.7 served as the basis for the coding form, consisting of five major categories: listening comprehension, reading comprehension, visual comprehension, integration of information, and listening/viewing strategies. Next, participants' comments were segmented into chunks that corresponded to the components listed in the coding scheme. Two additional codes emerged from the data: note viewing and processing of social cues. Task-specific strategies also emerged as a subcode of listening/viewing strategies. A description of the five major coding categories and their subcodes is presented in Table 12.

Based on the coding scheme, all chunks were coded using the NVivo 12 software (QSR International, 2018) and produced frequency counts for each category. For instance, the following comment from a participant who performed the task once was segmented into two chunks.

// [Chunk 1] "Then she talked about 'receptors'. I did not understand what this word meant, so I gave up and skipped it. // [Chunk 2] Then she mentioned two keywords of the lecture, 'rods' and 'cones'. I just wrote them down." //

"然后她又讲了 receptors，我没想明白这个词什么意思，我就略过放弃了。然后她提到了这个课的两个关键词，rods 和 cones，我就写了下来。"

Chunk 1 shows that the participant omitted a word from processing for comprehension, so it was categorized as *skipping*. Chunk 2 indicates that the participant recognized individual words in speech and was classified as *lexical search*. More examples of stimulated recall comments regarding learners' multimodal processing behaviors are presented in Table 12. It should be noted here that except for comments related to noticing the pronunciation of words, the data did not yield any direct evidence in support of the subcode input decoding. Other studies (e.g., Holzknecht, 2019; Rukthong & Brunfaut, 2020) have also reported that it was difficult to isolate evidence of input decoding from verbal reports because of its highly automated nature. Although participants did not specifically mention recognizing incoming sounds as speech, it was reasonable to assume that they would have relied on this process at a fundamental level during lecture viewing, as input decoding underlies all other cognitive processes (Field, 2013). Therefore, no example was given to illustrate this subcode. Furthermore, some chunks were only broadly classified as lower-level processes, given the limited amount of information provided (e.g., "I was writing down the keywords while listening"). In comparison to the lower-level processes, it was more straightforward to categorize higher-level processes into meaning construction and discourse construction, thus no examples were provided for the subcode of higher-level processes.

Table 12 Coding Scheme of Participants' Cognitive Processes during Lecture Viewing

Code	Description	Example quote (Translated from Mandarin)
Listening comprehension		
Lower-level processes	Encoding a message into language, including three operations: input decoding, lexical search, and parsing.	"I was writing down 'dendrites', 'arbor', and 'keep branching'."
Input decoding	Recognizing incoming sounds as speech.	N/A
Lexical search	Recognizing individual words in spoken form.	"I was thinking about the meaning of 'neuron'."
Parsing	Putting individual words into a syntactic pattern to form the bare meaning of an utterance at clause or sentence level.	"I wrote down 'meninges'. I wrote 'go from pia to dura on the far side'."
Higher-level processes	Extending the bare meaning of an incoming utterance, including two operations: meaning construction and discourse construction.	N/A
Meaning construction	Relating the literal meaning of utterances to the context in which they occurred to construct higher-level meaning.	"This thing (synapse) is responsible for 'transfer', which takes the form of 'conversation'. I was thinking this was her metaphor."
Discourse construction	Relating the meaning of the text to the discourse as a whole.	"At that time, I knew she was talking about a 'neuron' had four parts, and then she talked about the central part, the part with the DNA."
Reading comprehension (Graphemic input analysis)	Recognizing individual words in written form.	"The second time I noticed those two words she wrote, 'rods' and 'cones'."
Visual Comprehension		
Visual feature analysis	Identifying graphic displays in the picture.	"When I was looking at the diagram below, I was

Code	Description	Example quote (Translated from Mandarin)
		thinking this is the gap between cells."
Depictive processing	Mapping the graphic displays to form a mental model (i.e., representations of the key parts of the presented material and their relations).	"During the first viewing, I did not copy the picture completely because I did not have enough time. So, the second time I wanted to see 'four parts'. Yes, the logical relationship between the 'four parts'."
Integration of information		
Perceptual processing	Connecting words (spoken or written) and graphical elements.	"I was thinking about whether 'synapse' was referring to the gap in the middle or referring to the whole process. "
Semantic processing	Connecting propositional descriptions with structural characteristics of the mental model.	"I was drawing arrows following her. She then explained where the information came into the dendrites and cell body."
Listening/viewing strategies		
Cognitive		
Skipping	Ignoring a piece of information that is considered unimportant or unreliable.	"Although I did not know the meaning of these words, I did not think about what they meant, as long as I could understand its function."
Linguistic inferencing	Using information in the text to guess the meanings of unfamiliar words or fill in missing information.	"Here, I did not understand the word. She said there were three layers or something, and then she said 'pia' was very what, and 'dura' was 'very tough'. I was wondering if the first one was very fragile."
Kinesic inferencing	Using facial expressions, body language, and hand movements to guess the meaning of unknown words.	"Here, I was trying to understand the function of 'dura' through her body movements."
Elaboration	Embellishing interpretations of the text using prior	"Here, I visualized the scotopic condition based on what

Code	Description	Example quote (Translated from Mandarin)
	knowledge and relating them to knowledge gained from the text.	she said.”
Prediction	Anticipating what is going to hear using information from the text and prior knowledge.	“I thought she was going to talk about the middle one, but she did not, so I looked there for a while.”
Contextualization	Placing language items into a specific context in order to assist comprehension.	“I was thinking about whether the ‘dendrite’ was the vocabulary she introduced previously, and I realized that it actually was, so I added it to my notes.”
Reorganizing	Transferring processed information into forms that facilitate understanding, storage, and retrieval.	“I was thinking that I had to take notes, and my notes should not be messy. So, I wrote the four neuron parts following the order she introduced them.”
Translation	Relying on one’s knowledge of the L1 to facilitate comprehension.	“I was wondering if I had learned this before. I was recalling its Chinese name, but I could not.”
Metacognitive		
Planning	Developing a plan to overcome difficulties that may interfere with task completion.	“I knew she was going to introduce some functions of ‘cell body’, which I did not fully understand during the first viewing, so I was getting ready to listen to that feature.”
Focusing attention	Avoiding distractions and paying attention to the input.	“The first time I got very little information. She just briefly mentioned it. I only knew it (arachnoid) was a noun and a part of layers, so here I was focusing on listening to her.”
Monitoring	Checking, verifying, or correcting comprehension while performing the task.	“I was checking whether the ‘cell body’ was the same as ‘soma’. And whether its key information was correct.”
Evaluation	Checking the outcomes of comprehension based on an internal or external measure of completeness, reasonableness, and accuracy.	“The first time I listened to this part, I drew what it looked like, but I only wrote ‘half a Christmas tree’. The second time I wrote a ‘paddle’. The third time I thought

Code	Description	Example quote (Translated from Mandarin)
		I indeed needed to write it down."
Socio-affective (Managing emotions)	Being aware of and constructively handling negative emotions.	"I did not quite understand the previous part, but I did not need to worry about it since I could understand what she was saying in this part. So, I just started listening again."
Task-specific strategies	Goal-directed mental actions related to successful completion of the free recall test.	
Selecting information		"I was thinking about whether I should tell her the information that there were many 'units' between these two conditions."
Remembering pronunciation		"I was thinking that I was going to tell the classmate about this, so I thought I needed to memorize its (dendrite) pronunciation."
Remembering exact words		"I was memorizing the language she used so that I could describe it more properly. This was because what she said was more precise."
Note-viewing	Reading one's note.	"I was reading my notes, so I stopped listening to her."
Processing of social cues	Looking at the instructor's nonverbal communication cues, such as facial expressions and gestures.	"She was moving her hands, so my eyes followed her movements."

To answer RQ6 (To what extent does repeating a multimodal lecture-viewing task affect learners' awareness of technical words, as reflected in their stimulated recall comments?), only comments regarding the target words were considered. Target-word-related comments were further segmented into chunks, which were defined based on a coding scheme drawing on Schmidt's (1995, 2001) Noticing Hypothesis and Hegarty's (2014) information processing model. The initial coding scheme included two major categories: (a) level of awareness and (b) source of awareness. Noticing as the lower level of awareness referred to conscious registration of the item-level properties of a word. Four subcodes of noticing emerged from the data, including noticing orthographic form, phonological form, grapheme and phoneme correspondence (GPC), and POS. Understanding as the higher level of awareness involved knowing the meaning of a word. The source of awareness initially included subcategories of aural commentaries, diagram labels, diagram illustrations, and integrated information from different modalities. Two further subcodes, participants' notes and non-verbal signals, were added.

Based on the coding scheme developed for the investigation of awareness and vocabulary learning, all target-word-related chunks were coded and counted for each category using the NVivo 12 software (QSR International, 2018). For example, the following comment from a stimulated recall participant who performed the task twice was segmented into two chunks.

// [Chunk 1] "I was familiar with this diagram and I knew the structure of the system after the first viewing. So, the first thing I did during the second viewing was to check whether I spelled 'soma' correctly. // [Chunk 2] I was also thinking that I would leave a message to my classmate after viewing the lecture, so I wanted to make sure that I did not get anything wrong. I was checking whether the cell body was the same as "soma." //

“看完一遍之后我已经知道它这个图，和它这个系统的结构，所以我第二遍首先第一个干的事情就是我要确认那个 soma 拼对没。因为我当时也想的是要看完之后给同学留这个 message, 然后我在确认它的关键信息有没有错， 我在确认 cell body 和 soma 它是同一个东西。”

Chunk 1 shows that the participant was paying attention to the spelling of a target word, so it was coded as *noticing*, *from diagram labels*. Chunk 2 indicates that the participant was confirming the meaning of a target word based on the instructor's spoken input. Therefore, it was categorized as *understanding*, *from aural commentaries*. For more examples of stimulated recall comments regarding participants' awareness of target words, see Table 13.

Table 13 Examples of Stimulated Recall Comments Regarding Participants' Awareness of Target Words

Code	Example quote (Translated from Mandarin)
Level of awareness	
Noticing	
Not specified	"I was focusing on writing down these three words (scotopic, mesopic, and photopic)."
Orthographic form	"I was correcting the spelling of the word 'axon'. The first time I misspelled it as 'axom', but I noticed that the letter was actually 'N' during the second viewing."
Phonological form	"I was thinking about the pronunciation of this word (dendrite). It was pronounced /aɪ/. As I knew I was going to tell the classmate about this, I thought I should remember its pronunciation."
GPC	"I was feeling confused because she was pointing at 'scotopic' while saying 'photon'. I felt strange because I could not match the two pieces of information."
POS	"The word 'meningeal' was an adjective and then changed to a noun (meninges) in the next sentence."
Understanding	"For this part, I heard her loud and clear. She said that dura floated our brain in fluid and prevented us from having concussions."

Code	Example quote (Translated from Mandarin)
Source of awareness	
Not specified	"I wrote down the word 'meninges'."
Aural commentaries	"While she was introducing its feature (scotopic), I added two more points, which were 'dark condition' and 'not much light'."
Diagram labels	"I thought she was going to talk about the middle one (mesopic), but she did not. So, I looked at the middle one for a while."
Diagram illustrations	"What she drew (synapse) looked like a nose, so I thought the information was going to be transferred to the nose."
Multiple input sources	"I was thinking about what this black circle represented, and then I realized it represented the scotopic condition she mentioned, which was very dark."
Notes	"I knew this was the end of the video, and there was no more valuable information, so I was reading my notes while memorizing the structure of these neuron parts (soma, dendrites, axon, and synapse)."
Nonverbal signals	"While viewing this part for the second time, I was trying to understand the function of dura through her body movements."

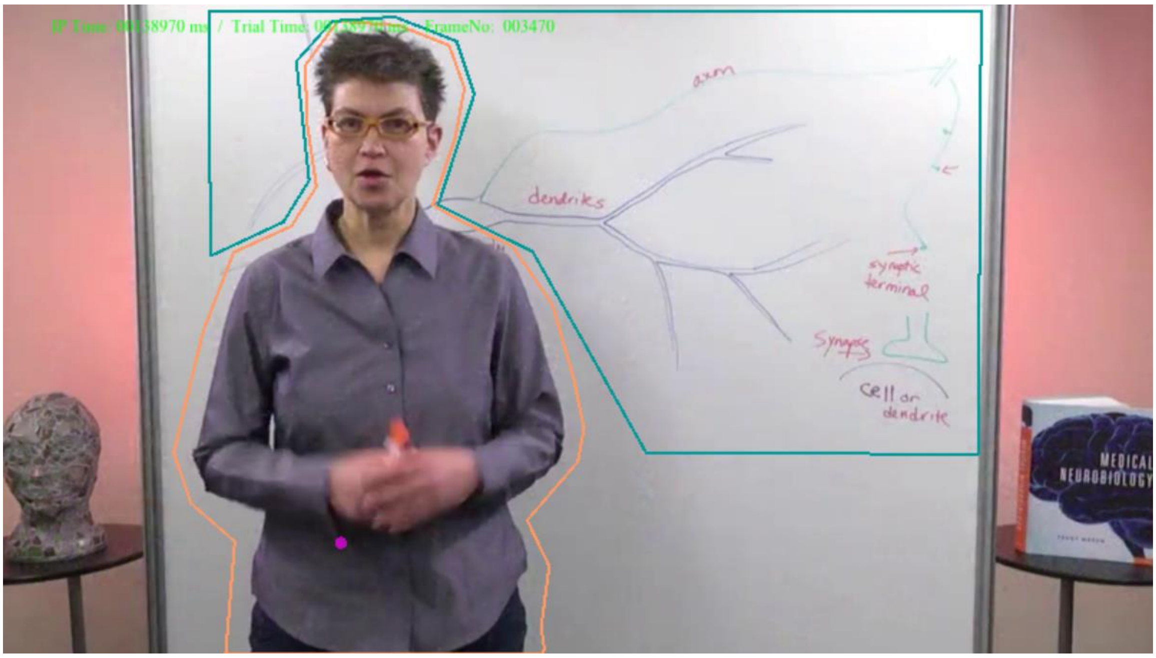
Finally, 20 percent of the data were coded by a second coder, a Mandarin speaker of L2 English with an L2 research background, using the same coding schemes. The inter-coder reliability was high (Cohen's kappa = .81). Appendix 11 includes exemplary excerpts from the stimulated recalls.

3.7.6 Eye-tracking Data

The eye-tracking data were analyzed with EyeLink Data Viewer (2019) software. Data cleaning was performed before analyses, following recommendations in previous research (Conklin et al., 2018; Godfroid, 2020). Eye-movement recordings were first inspected in a trial-by-trial manner to identify low-quality data. Three participants' data were excluded from analysis as their data contained abnormal track loss, demonstrated in the temporal graph and spatial overlay view by plotting the raw data in the software. Drift correction was then performed for 26 out of 120 trials. All fixations in a problematic trial were moved up or down manually at once, as the identified drift indicated a systematic offset between the recorded eye gaze location and a participant's true eye gaze location. Fixations shorter than 80 milliseconds were removed from the dataset (8.81% of the data), and other fixations were not merged. The next step was to select interest periods for each video. Parts that were not directly relevant to the lecture were excluded from the interest periods, such as video openings (i.e., still images presenting information about the lecture) and closings (i.e., previews of the next video). A two-minute course warm-up at the beginning of the first video was also excluded, given that it only aimed at familiarizing learners with the topic and preparing them for subject learning. The interest periods of the three videos were 196461, 317919, and 380111 milliseconds, respectively.

To investigate the effect of task repetition on learners' visual attention to the instructor and the diagrams (RQ1) and the relationship between visual attention to these areas and lecture comprehension (RQ2), I created dynamic AOIs by manually drawing irregular areas around the instructor and diagrams (see Figure 20). Unlike static interest areas which represent fixed areas of a visual stimulus that remain constant throughout a trial, dynamic interest areas can be defined as a series of instances of static interest areas, with each instance having both an onset time and an offset time. This enables dynamic interest areas to change position, size, and shape as the trial progresses. By utilizing dynamic AOIs in this manner, I was able to capture the movements of the instructor and diagrams across different time segments. For example, the instructor appeared on the screen at the beginning of the video for 100 milliseconds. She then started to move to the left of the screen at time 101 milliseconds and stopped at time 168 milliseconds. This required the creation of three instances to represent the instructor's changing position (see Table 14). The first instance was activated at time 0 millisecond and deactivated at time 100, as the instructor did not move within this time period. Next, the other two instances were created to align over the instructor's moving position frame by frame, with each being activated for 33 milliseconds. This is also the minimum length of an instance (one video frame lasted approximately 33 milliseconds). Considering that the instructor's movements resulted in changes in diagram size, the duration of the two AOIs was set the same. To account for potential variations in attention allocation among the participants due to differences in instance-level factors (i.e., instance size and duration), the final statistical analyses included these differences as covariates. Descriptive statistics for the instructor and the diagram AOIs are presented in Table 15. More examples of instances created for the AOIs are demonstrated in Figures 21 and 22.

Figure 20 An Example of the Instructor and the Diagram AOIs (Enclosed in Yellow and Green Irregular Shapes)



Note. Instance 1 of the instructor AOI: size = 510082 pixels; duration = 1057 milliseconds; Instance 1 of the diagram AOI: size = 554857 pixels; duration = 1057 milliseconds.

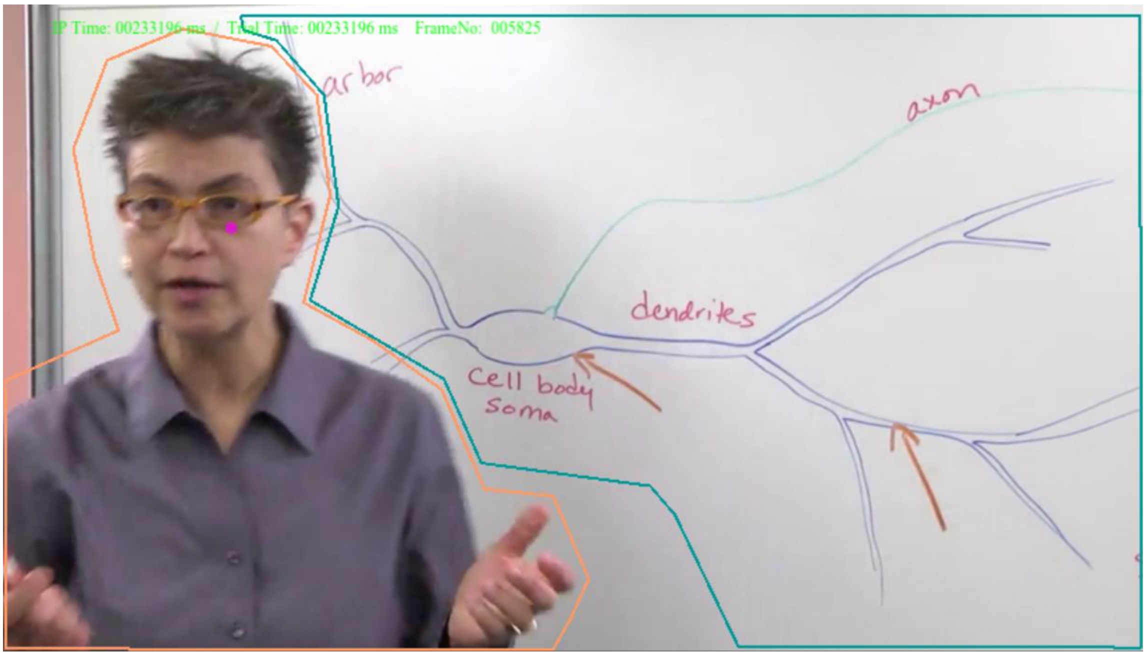
Table 14 An Example of an Instance List

Instance Number	Onset time (milliseconds)	Offset time (milliseconds)	Duration (milliseconds)
1	0	100	100
2	101	134	33
3	135	168	33

Table 15 Descriptive Statistics for the Instructor and the Diagram AOIs

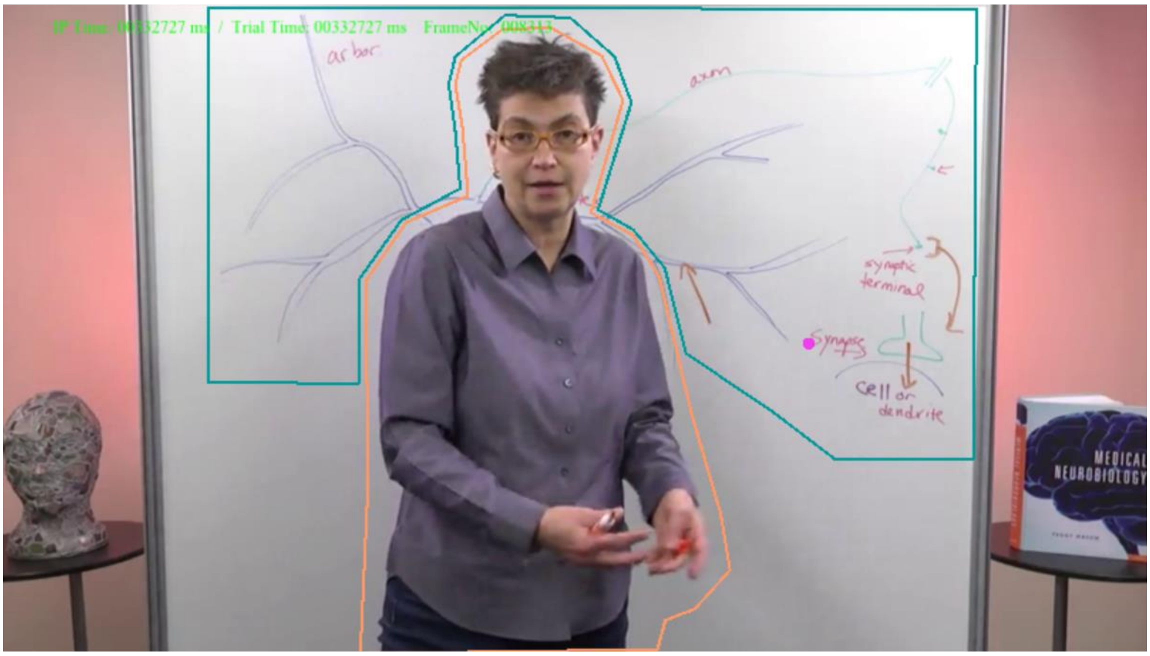
AOI	Number of instances	Size (pixels)		Duration (milliseconds)	
		M (*SD*)	95% CI	*M* (*SD*)	95% CI
Instructor	482	603193 (133033)	[591287, 615100]	407. 60 (475.88)	[365.00, 450.19]
Diagram	482	826714 (270489)	[802506, 850923]	407. 60 (475.88)	[365.00, 450.19]

Figure 21 An Example of the Instructor and the Diagram AOIs When Video Zoomed In



Note. Instance 244 of the instructor AOI: size = 652809 pixels; duration = 198 milliseconds; Instance 244 of the diagram AOI: size = 1150157 pixels; duration = 198 milliseconds.

Figure 22 An Example of the Instructor and the Diagram AOIs When Video Zoomed Out



Note. Instance 480 of the instructor AOI: size = 488329 pixels; duration = 694 milliseconds; Instance 480 of the diagram AOI: size = 571718 pixels; duration = 694 milliseconds.

Initially, three eye-movement measures were extracted for each AOI from individual viewings, including total fixation duration, fixation count, and integrative saccades (see Table 16 for description). The two aggregate late eye-movement measures (i.e., total fixation duration/count) were assumed to provide a general picture of participants' attention allocation to the target words (Godfroid, 2020). The number of transitions between AOIs has been argued to be indicative of integration of information from different sources (e.g., Arndt et al., 2015; Hegarty & Just, 1993; Johnson & Mayer, 2012; Scheiter & Eitel, 2017). To examine how task repetition affected the relationship between attention allocation and comprehension, cumulative fixation duration was also calculated by adding up the total fixation duration on the instructor/diagram AOIs during each individual viewing.

All continuous eye-tracking metrics (i.e., fixation duration and cumulative fixation duration) were log-transformed in advance to meet the normal distribution assumption. As the current study included data with a nested structure (i.e., each participant contributed multiple observations), the utilization of mixed-effects models would be an optimal approach to investigate the effects of task repetition while taking into consideration the random variation that might exist within and between individual participants and instances of the AOIs (further description of this approach is presented in Section 3.8). During the instance-level analysis, it was observed that a large proportion of the instructor and the diagram AOIs were skipped (i.e., recorded as 0 millisecond). Even after logarithmic transformation, this presence of a high number of zeros could potentially lead to violations of the statistical assumptions underlying mixed-effects regression. Thus, skipped instances were excluded from the fixation duration and count analyses, and the measure of AOI instance skip rate (i.e., the proportion of instances of a dynamic AOI that were not fixated upon) was included. For example, in a dynamic AOI consisting of 5 instances, if a participant looked at only 1 of the instances, then the instance skip rate would be calculated as the number of skipped instances divided by the total number of instances, which is 4/5 in this case.

Furthermore, I had originally intended to extract the measure of mean fixation duration, as it might provide additional insights into the level of cognitive processing demands and complexity of the task. However, in conducting mixed-effects regression, I found that this measure was not suitable because the majority of instances only contained a single fixation, and so the mean fixation duration would closely approximate to total fixation duration when doing instance-level analysis. This made the metric less informative for the purpose of the analysis and thus was not included in the final statistical analyses. It is also noteworthy that despite my initial plan to analyze eye-movement measures extracted from all three videos, I ultimately decided to only focus on a subset of the dataset, specifically the data extracted from the first video, due to the complexity of manually creating dynamic AOIs.

Table 16 Description of Eye-Movement Measures

Eye-movement measure	Description
Total fixation duration	The sum of all fixation duration on an AOI
Mean fixation duration	Total fixation duration on an AOI divided by the number of fixations on the AOI
Fixation count	Total number of fixations on an AOI
Run count	Total number of times an AOI was entered and left
Integrative saccades	Number of integrative transitions between the instructor and the diagram AOIs
Cumulative fixation duration	The sum of all fixation duration on an AOI during the entire treatment
Mean cumulative fixation duration	Cumulative fixation duration on an AOI divided by the total number of fixations on the AOI during the whole treatment
AOI instance skip rate	The proportion of instances of a dynamic AOI that were not fixated upon
AOI skip rate	Whether an AOI was fixated upon: a target word AOI was considered skipped if no fixation was fixated upon

The correlations between different eye-movement measures extracted for the instructor and the diagram AOIs were then computed. The results are presented in Table 17, showing strong positive correlations between total fixation duration and fixation count on the two AOIs. This indicated that duration and count measures might be assessing similar aspects of visual attention. A strong positive correlation was also revealed between the number of integrative saccades and three measures: fixation duration on the instructor AOI, fixation count on the instructor AOI, and fixation count on the diagram AOI. This might suggest that increased visual attention to the AOIs is related to more integrative transitions between these AOIs. Finally, strong negative correlations were found between instance skip rate and variables of fixation duration, fixation count, and the number of integrative saccades, respectively, indicating that it is warranted to include the measure of instance skip rate in conjunction with duration and count measures.

Table 17 A Correlation Matrix of the Eye-Movement Indices Extracted for the Instructor and the Diagram AOIs

Eye-movement measure	1	2	3	4	5	6	7
1. Fixation duration on the instructor AOI	—	.48	.93	.47	.82	-.96	-.61
2. Fixation duration on the diagram AOI		—	.48	.92	.62	-.53	-.95
3. Fixation count on the instructor AOI			—	.56	.90	-.96	-.64
4. Fixation count on the diagram AOI				—	.71	-.55	-.94
5. Number of integrative saccades					—	-.89	-.76
6. Instance skip rate of the instructor AOI						—	.66
7. Instance skip rate of the diagram AOI							—

Note. all $p < .001$

To investigate the effect of task repetition on learners' visual attention to the technical words (RQ5) and on the relationship between visual attention and vocabulary gains (RQ7), 11 dynamic AOIs were created for each occurrence of the target words in the videos. As demonstrated in Figures 23 and 24, the AOIs were adjusted to capture the changes in size due to camera zooming in and out. They were only activated during the time the target words were presented. Since each target word AOI had a limited range of size variations (up to three), a weighted mean size was calculated for the AOIs, taking into consideration the duration of each AOI size appearing on the screen. For instance, the AOI created for the word *synapse* had a small size (13020 pixels) for 120421 milliseconds and a large size (32265 pixels) for 36383 milliseconds. The time weight for the small size was 76.80 % (dividing the presentation time of the small size by the total presentation time: 120421 / 152686) and the weight for the large size was 23.20%

(dividing the presentation time of the large size by the total presentation time: 32265 / 152686). Therefore, the weighted mean size of this AOI was 17485 pixels, obtained by multiplying each size value by its time weight and taking the sum (13020 * 76.80% + 32265 * 23.20%). The weighted mean size and presentation duration of each target word are listed in Table 18. These item-level differences were also included as covariates in the statistical analyses.

Figure 23 An Example of the Target Word AOIs (Enclosed in Rectangles)

Figure 24 An Example of the Target Word AOIs When Video Zoomed In

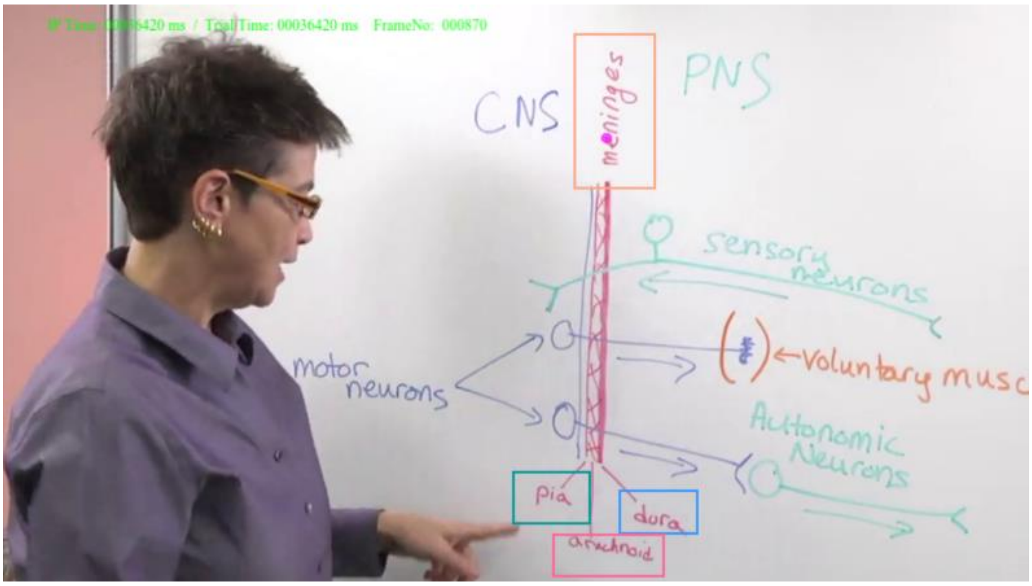


Table 18 Descriptive Statistics for the Target Word AOIs (in Alphabetical Order)

Item	Presentation duration (milliseconds)	Weighted mean AOI[a] size (pixels)
arachnoid	143.91	11433
axon	186.45	15097
dendrite	183.38	23954
dura	190.99	7157
meninges	291.77	25088
mesopic	256.08	17577
photopic	285.20	22018
pia	121.66	9332
scotopic	275.76	17577
soma	127.64	12869
synapse	156.80	17485

Note. [a]Area of Interest

For individual viewings, five eye-movement measures were extracted for each target word AOI, including total fixation duration, mean fixation duration, fixation count, run count, and AOI skip rate for individual viewings (see Table 12 for a description of measures). Again, cumulative fixation duration was manually calculated as the sum of total fixation duration on target word AOIs in each viewing. All continuous eye-tracking metrics (i.e., fixation duration, mean fixation duration, and cumulative fixation duration) were log-transformed in advance, and mixed-effects modeling was employed to account for the random variation within and between individual participants and target items. Results of the Pearson correlations for the eye-movement measures extracted for the target word AOIs are presented in Table 19. The results revealed strong positive correlations between total fixation duration and mean fixation duration, as well as between fixation count and run count, suggesting that participants tended to fixate upon the target word AOIs consistently over time and made multiple entries and exits from the target word AOIs when they fixated on them. The positive correlations also indicated that duration and count measures were gauging related aspects of visual attention, respectively. The strong negative correlations between AOI skip rate and both total fixation duration and fixation count indicated that AOI skip rate was also a reliable measure of eye movements.

Table 19 A Correlation Matrix of Eye-Movement Indices Extracted for the Target Word AOIs

Eye-movement measure	1	2	3	4	5
1. Total fixation duration	—	.86	.66	.66	-.80
2. Mean fixation duration		—	.27	.28	-.95
3. Fixation count			—	.96	.23
4. Run count				—	.24
5. AOI skip rate					—

Note. all $p < .001$

3.8 Statistical Analyses

The descriptive and correlational statistical analyses were conducted with SPSS (Statistical Package for the Social Sciences) version 27.0. Intercorrelations between various eye-movement indices and between different test scores were computed using Pearson’s correlations. The reliability of the vocabulary post-tests was established using Cronbach’s alpha. The intercoder reliability of the stimulated recall, free recall, and meaning recall tests was measured by Cohen’s kappa. Regression models were constructed with the *lm* function of the *stats* package. A series of mixed-effects models, that is, models containing both fixed effects (independent variables) and random effects (participants and items), were also constructed using the R software (R Core Team, 2019) with the *lmer* and *glmer* function of the *lme4* package (Bates et al., 2015). The alpha level was set at .05 for all tests.

There are several advantages of using mixed-effects modeling over standard regression analyses. First, it allows for "the simultaneous generalization of the results on new items and new participants" as it treats participants and items as random variables in one model (Gagné & Spalding, 2009, p. 25). Second, it enables examining main independent variables in addition to other participant-level and item-level covariates, leading to increased precision of the estimates of the fixed and random effects (Baayen et al., 2008; Cunnings, 2012; Linck & Cunnings, 2015). Also, it can cope with different types of dependent variables (continuous, binary, or count data) through linear and generalized linear mixed-effects models. Lastly, it can handle missing values and imbalanced designs (Linck & Cunnings, 2015; Sonbul & Schmitt, 2012). Considering that the current study included continuous (e.g., fixation duration), binary (e.g., correct/incorrect answers), and count (e.g., fixation count) data as well as various participant- and item-related covariates, mixed-effects modeling was considered appropriate for data analyses.

For the first research question, four eye-movement metrics (total fixation duration, fixation count, skip rate within the instructor and the diagram AOIs, as well as the number of integrative saccades between the two AOIs) were the dependent variables, and time (first, second, and third viewing) was set as the fixed effect. For the third research question, I investigated the relationship between the control group's free recall scores (dependent variable), as measured by the number of correct and incorrect idea units, and their total fixation duration on the instructor and the diagram AOIs (predictors). I also examined the associations of the repetition group's free recall scores (dependent variable) with their total fixation duration on the same AOIs at each exposure and during the entire treatment (predictors). For the third research question, frequency of counts was calculated for each code under different conditions (performing the task once, twice, or three times). To address the fourth research question, in separate analyses for the vocabulary tests (immediate and delayed form recognition, meaning recall, and meaning recognition), the dependent variables were the post-test scores, while group (control versus repetition) was set as the fixed effect. To answer the fifth research question, the eye-movement metrics (total fixation duration, mean fixation duration, fixation count, run count, and skip rate within the target word AOIs) were the dependent variables, and time was set as the fixed effect. For the last research question, I looked at the relationship between total fixation duration and the control group's vocabulary post-test scores, as well as between the cumulative eye-movement metric (the sum of total fixation duration on the target word AOI during each viewing) and the repetition group's vocabulary post-test scores. Detailed information about model construction is provided in the next chapter.

For each analysis, the modeling started by constructing a null model that contained only random intercepts. Linear mixed-effects models were built for continuous dependent variables using the *lmer* function, Poisson mixed-effects models were constructed for count dependent variables using the *glmer* function with the argument *family = Poisson*, and logistic mixed-effects models were fit for the binary dependent variable using the *glmer* function with the argument *family = binomial*. Fixed effects were then entered into the null model step-wise and tested using likelihood ratio tests to check whether the inclusion of the fixed effects significantly improved model fit. After identifying the fixed effects, model comparisons were carried out using a backward model selection approach to arrive at the best-fit model (Barr et al., 2013). I started with the maximal random effects structure and progressively trimmed any random effects that did not significantly improve model fit (Matuschek et al., 2017). Changes in model fit were measured by likelihood ratio tests as well as absolute Akaike Information Criterion (AIC) values. When models with maximal random structure failed to converge, the first step was to add an optimizing function using different control augments to refit the model

(Linck & Cunnings, 2015): *control = lmerControl(optimizer ="Nelder_Mead")* for linear mixed-effects models and *control = glmerControl(optimizer = "bobyqa")* for generalized linear mixed-effects models. If this did not resolve the convergence issue, random effect parameters resulting in the least variance were removed one by one until convergence was achieved (Blom et al., 2012).

Participant-level variables (CAE and VST scores) and item-level variables (item length, number of syllables, POS, frequency of occurrences in the spoken text, number of nonverbal signals, number of elaborations, AOI presentation time, and AOI size) were then entered into the models as categorical or continuous covariates. Continuous covariates (VST scores, AOI presentation time, and AOI size) were log-transformed in advance to ensure the covariates were on the same scale. Some continuous covariates were also centered (i.e., subtracting the mean value of the predictor from each individual value) to solve convergence issues. Covariates were only kept in the model when they contributed significantly to the model fit. Finally, predictors were removed backward to check for predictors not yielding significant differences, but none was found to be redundant.

The collinearity, normal distribution of residuals, and homoscedasticity assumptions were checked for all linear mixed-effects models using the *sjPlot* package (Lüdecke, 2022). The multicollinearity assumption was checked for all logistic mixed-effects models using the *car* package (Fox & Weisberg, 2019), and the linearity of continuous predictors (i.e., fixation duration and mean fixation duration) and the logic of the outcome variables (the vocabulary post-test scores) were also examined. The assumption for Poisson mixed-effects models was that the mean of a Poisson random variable must be equal to its variance, and this was checked using the *performance* package (Lüdecke et al., 2021). Overdispersion was detected in two Poisson mixed-effects models. Therefore, observation-level random effects were used to incorporate overdispersion, that is, adding an *obs_effect* variable with a unique value for each observation to the models (Harrison, 2014).

After all models were fit and model diagnostics were carried out to ensure the data met the assumptions, outliers were identified in linear mixed-effects models as those data points with absolute standardized residuals exceeding 3 standard deviation (SD, $3 < |z|$) using the *LMERConvenienceFunctions* package (Tremblay & Ransijn, 2020). Influential points in generalized mixed-effects models were detected using the *cooks.distance()* function. The cut-off value for Cook's distance was set at $4/n$ where n refers to the number of groups in the grouping factor. After outliers were excluded, the same models were refit and the results were compared with the original analysis. Outliers did not have any effects on the model results and therefore were kept in the dataset. Effect sizes for fixed effects (marginal R^2 and conditional R^2) were obtained using the *r.squaredGLMM* function in the *MuMln* package (Bartoń, 2022). R-squared was used to quantify effect sizes in the linear regressions. Odds ratio was used as an alternative applicable to logistics regression to measure effect size (Field et al., 2012), and was considered strong when greater than 3 or less than 0.33 (Haddock et al., 1998).

Chapter 4 Results

This chapter describes the results obtained from the quantitative and qualitative analyses. First, the results of preliminary analyses are presented to ensure the reliability of the instruments and thus the validity of the research. In particular, I considered the reliability of immediate and delayed vocabulary post-tests, the equivalence of proficiency level between groups, participants’ prior knowledge of the target items, and potential effects of topic familiarity on their lecture comprehension and vocabulary acquisition. Next, this chapter explores how task repetition influenced participants’ visual attention to the instructor and the diagram during lecture viewing and reports the findings on the participants’ cognitive processes and strategies activated to understand the multimodal input. This is followed by the examination of the relationship between learners’ processing of visual stimuli (i.e., the instructor and the diagram) and lecture comprehension. The chapter then investigates the effect of task repetition on the acquisition of technical words, as well as the way in which participants allocated visual attention to the target words across repeated viewing. Then, the chapter reports participants’ awareness of the target words and the association between their visual attention to the target words and vocabulary gains.

4.1 Preliminary Analyses

4.1.1 Reliability of Vocabulary Post-Tests

The internal consistency of immediate and delayed form recognition and meaning recognition tests were examined via Cronbach’s alpha. As presented in Table 20, the immediate and delayed form recognition tests had an α of .52 and .54, respectively, and the immediate and delayed meaning recognition tests had an α of .67 and .64, respectively. The values of Cronbach’s alpha were low for the tests, which might be due to the small number of items included. In addition, the mean score for the immediate and delayed form recognition tests was quite high, implying a potential ceiling effect, which might have contributed to the low reliability coefficient as well.

Table 20 Descriptive Statistics for Immediate and Delayed Form Recognition and Meaning Recognition Test Scores

Vocabulary test	*M (SD)*	95% CI	Cronbach's alpha
Immediate form recognition	17.50 (2.55)	[16.84, 18.16]	.52
Delayed form recognition	17.72 (2.41)	[17.09, 18.34]	.54
Immediate meaning recognition	7.57 (2.39)	[6.95, 8.18]	.67
Delayed form recognition	7.02 (2.38)	[6.40, 7.63]	.64

Note. Maximum score for: immediate and delayed form recognition = 22; immediate and delayed meaning recognition = 11.

4.1.2 Equivalence between Groups

Descriptive statistics for the participants' proficiency test results (i.e., CAE and VST) are presented in Table 6. To ensure that the control and the repetition groups were at an equivalent proficiency level, linear regression models were constructed without including any random effects, given that each participant had only one overall comprehension score and one vocabulary size score. The results confirmed that these two groups did not differ from each other in terms of their listening proficiency, $R^2 < .001$, $F(1, 58) = 0.04$, $p = .84$, $d = .02$, or vocabulary size, $R^2 < .001$, $F(1, 58) = 0.003$, $p = .96$, $d = .03$.

4.1.3 Prior Knowledge of Target Words

All participants were asked to indicate their prior knowledge of target words on an unfamiliar-familiar continuum (1 = "I didn't know this word before watching the video"; 5 = "I definitely knew this word before watching the video") on the post-experiment questionnaire. The questionnaire confirmed that the participants had little prior knowledge of the target words (for descriptive statistics, see Table 21). To check whether the control and the repetition groups' prior knowledge of the target words had any effects on their vocabulary gains, a series of logistic mixed-effects models were constructed with participants' prior knowledge of the target word as the fixed effect, and subject and item as the random effects. The outcome variables were immediate and delayed form recognition, meaning recognition, and meaning recall test scores, respectively. Likelihood ratio tests were conducted to compare null models with subject and item as random effects and models including target word familiarity as the fixed effect. The results showed that the inclusion of topic familiarity did not make a significant difference to the null models (immediate form recognition, $\chi^2(1) = .03$, $p = .87$, $R^2 < .01$; immediate meaning recognition, $\chi^2(1) = .85$, $p = .36$, $R^2 < .01$; immediate meaning recall, $\chi^2(1) = .85$, $p = .36$, $R^2 < .001$; delayed form recognition, $\chi^2(1) = 0.52$, $p = .28$, $R^2 = .01$; delayed meaning recognition $\chi^2(1) = .25$, $p = .62$, $R^2 < .001$; delayed meaning recall, $\chi^2(1) = .23$, $p = .64$, $R^2 < .001$, suggesting that the two groups' knowledge of target words prior to the experiment did not affect their vocabulary gains.

Table 21 Descriptive Statistics for Participants' Prior Knowledge of the Target Words by Item (in Alphabetic Order)

Target item	*M (SD)*	95% CI
arachnoid	1.05 (0.22)	[0.99, 1.11]
axon	1.18 (0.57)	[1.04, 1.33]
dendrite	1.10 (0.40)	[1.00, 1.20]
dura	1.10 (0.35)	[1.01, 1.19]
meninges	1.08 (0.38)	[0.98, 1.18]
mesopic	1.00 (0.00)	[1.00, 1.00]
photopic	1.17 (0.49)	[1.04, 1.29]
pia	1.15 (0.48)	[1.03, 1.27]
scotopic	1.03 (0.18)	[0.99, 1.08]
soma	1.05 (0.22)	[0.99, 1.11]
synapse	1.07(0.31)	[0.99, 1.15]

Note: Maximum value for each item = 5.

4.1.4 Topic Familiarity

Participants' familiarity with the topic was measured using a 5-point Likert scale question (i.e., "the topic of the lecture was familiar", ranging from "strongly disagree" to "strongly agree") included in the perception questionnaire. All participants indicated at least moderate unfamiliarity with the topic, *M* = 1.65, *SD* = 0.80, 95% CI [1.44, 1.86]. To examine the effect of topic familiarity on participants' lecture comprehension and vocabulary gains, Poisson and logistic mixed-effects models were constructed. While topic familiarity served as the fixed effect, subject and item were set as the random effects. The outcome variables were the number of correct idea units, immediate and delayed form recognition, meaning recognition, and meaning recall scores, respectively. When compared with a null model that contained only random effects, the results showed that including topic familiarity as the fixed effect did not make a significant difference in terms of the number of correct idea units, $\chi^2(1) = .98$, $p = .32$, $R^2 = .01$; immediate form recognition, $\chi^2(1) = .11$, $p = .74$, $R^2 < .001$; immediate meaning recognition, $\chi^2(1) = .31$, $p = .58$, $R^2 < .01$; immediate meaning recall, $\chi^2(1) = .15$, $p = .70$, $R^2 < .001$; delayed form recognition, $\chi^2(1) = 1.52$, $p = .22$, $R^2 = .02$; delayed meaning recognition, $\chi^2(1) = 1.19$, $p = .28$, $R^2 = .01$; and delayed meaning recall, $\chi^2(1) = .03$, $p = .87$, $R^2 < .001$. The results confirmed that the participants' topic familiarity did not affect their lecture comprehension or vocabulary gains.

4.1.5 Lecture Comprehension

Descriptive statistics for the number of idea units recalled by the participants are presented in Table 22, demonstrating that both the control and the repetition groups achieved adequate comprehension of the lecture. To check whether task repetition would affect the participants' lecture comprehension, Poisson mixed-effects models were built with group as the fixed effect, and subject and video as the random effects. The outcome variables included the number of correct and incorrect idea units, as well as the number of idea units that were not mentioned in the video transcripts. Compared with null models only containing the random effects, adding group as the fixed effect significantly improved the null models for the number of correct idea units, $\chi^2(1) = 30.44$, $p < .001$, $R^2 = .33$, and the number of incorrect idea units, $\chi^2(1) = 490.55$, $p < .001$, $R^2 = .06$, but not for the number of idea units that were not given, $\chi^2(1) = 1.28$, $p = .26$, $R^2 = .02$. As demonstrated in Table 23, the repetition group recalled significantly more correct and fewer incorrect idea units than the control group, but task repetition did not emerge as a significant predictor of the number of idea units that were not mentioned in the transcripts.

Table 22 Descriptive Statistics for the Number of Idea Units by Group

Group	Total number of idea units		Number of correct idea units		Number of incorrect idea units		Number of idea units that were not mentioned	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Video 1								
Control group (n =30)	24.27 (10.93)	[20.19, 28.35]	17.67 (9.27)	[14.20, 21.13]	3.10 (2.28)	[2.25, 3.95]	3.13 (2.70)	[2.13, 4.14]
Repetition group (n = 30)	36.43 (10.79)	[32.41, 40.46]	29.7 (9.79)	[26.11, 33.42]	1.83 (1.91)	[1.12, 2.54]	4.37 (3.12)	[3.20, 5.53]
Video 2								
Control group	26.37 (14.73)	[20.87, 31.87]	15.30 (8.99)	[11.94, 18.66]	4.17 (4.36)	[2.54, 5.79]	4.80 (6.56)	[2.35, 7.25]
Repetition group	42.10 (36.12)	[36.12, 48.08]	32.03 (13.03)	[27.17, 36.90,]	3.43 (4.37)	[1.80, 5.06]	4.97 (5.56)	[2.89, 7.04]
Video 3								
Control group	36.23 (15.89)	[30.30, 42.17]	24.23 (11.10)	[20.09, 28.38]	4.77 (4.34)	[3.15, 6.39]	3.60 (4.58)	[1.89, 5.31]
Repetition group	52.00 (18.24)	[45.19, 58.81]	42.03 (16.12)	[36.01, 48.05]	2.97 (3.20)	[1.77, 4.16]	4.47 (6.02)	[2.22, 6.72]
Total								
Control group	86.87 (37.66)	[72.80, 100.93]	57.20 (25.77)	[47.58, 66.82]	12.03 (7.96)	[9.06, 15.01]	11.53 (11.22)	[7.34, 15.72]
Repetition group	130.53 (41.40)	[115.00, 145.99]	103.83 (34.82)	[90.83, 116.83]	8.23 (7.43)	[5.46, 11.01]	13.80 (13.00)	[8.95, 18.65]

Table 23 Results of Between-Group Comparisons for Lecture Comprehension

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Number of correct idea units	0.62	[0.42, 0.82]	0.10	1.86	[1.53, 2.26]	6.28	**<.001**	.33	.87
Number of incorrect idea units	-0.44	[-0.83, -0.06]	0.19	0.64	[0.44, 0.93]	-2.32	**.02**	.06	.63
Number of idea units that were not given	0.26	[-0.19, 0.72]	0.23	1.29	[0.83, 2.02]	1.14	.26	.02	.74

Note. Values of *p* in boldface are significant at the .05 level of alpha.

4.2 RQ1: Processing of Multimodal Input during Repeated Viewing

To answer the first research question (To what extent does repeating a multimodal lecture-viewing task affect learners' visual attention to the lecture instructor and labeled diagrams, as reflected in their eye movements?), the repetition group's visual attention to the instructor and the diagram AOIs during each viewing of the first video was explored. Descriptive statistics for the eye-movement indices are provided in Table 24. It should be noted that the table presents the instance skip rate of the instructor and the diagram AOIs, that is, the proportion of instances of the AOI that were not fixated by participants (calculated as the number of skipped instances divided by the total number of instances).

Based on the type of dependent variables, various mixed-effects models were built with time set as the fixed effect. Linear mixed-effects models were constructed for continuous dependent variables (i.e., total fixation duration within the instructor and the diagram AOIs); Poisson mixed-effects models were constructed for count dependent variables (i.e., fixation count within the instructor and the diagram AOIs); and logistic mixed-effects models were built for binary dependent variables (i.e., skip rate within the instructor and the diagram AOIs). Subject and AOI instance (i.e., a freehand region drawn around the instructor or the diagram that was activated for a certain time period) were added as random intercepts. Some models also included by-subject random slopes for time. AOI instance size (i.e., pixel areas of a dynamic AOI at a particular point in time) and duration (i.e., the length of time that a dynamic AOI instance was active) were added to the models as covariates. In addition, a Poisson mixed-effects model was built to examine if time was a significant predictor of the number of integrative saccades between the instructor and the diagram AOIs. The model only contained by-subject intercepts as each participant's number of saccades was measured repeatedly. For best-fit model structures, see Table 25.

Table 24 Descriptive Statistics for the Eye-Movement Measures Extracted for the Instructor and the Diagram AOIs

AOI	Eye-movement measure	Time 1		Time2		Time 3	
		M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Instructor	Total fixation duration (milliseconds)	23249 (14772)	[17733, 28765]	19726 (12583)	[15027, 24424]	28131 (15135)	[22480, 33782]
	Fixation count	82.37 (40.84)	[67.12, 97.62]	68.20 (42.53)	[52.32, 84.08]	91.30 (50.42)	[72.47, 110.13]
	Instance skip rate	0.83 (0.38)	[0.82, 0.84]	0.86 (0.35)	[0.85, 0.86]	0.81 (0.39)	[0.80, 0.82]
Diagram	Total fixation duration (milliseconds)	49642 (16259)	[43571, 55713]	36596 (16578)	[30406, 42786]	37551 (18546)	[30626, 44476]
	Fixation count	192.23 (52.38)	[172.68, 211.79]	148.50 (64.84)	[124.29, 172.71]	150.03 (76.72)	[121.38, 178.68]
	Instance skip rate	0.59 (0.49)	[0.58, 0.59]	0.69 (0.46)	[0.68, 0.70]	0.66 (0.47)	[0.65, 0.67]
Instructor and diagram	Number of integrative saccades	75.67 (26.21)	[65.88, 85.46]	59.47 (24.88)	[50.18, 68.76]	75.03 (37.73)	[60.94, 89.12]

Table 25 Model Structures for the Instructor and the Diagram AOIs

AOI	Model structure
Instructor	FD[a] ~ Time + Size[b] + Duration[c] + (1 + Time | Subject) + (1 | Instance)
	FC[d] ~ Time + Size + Duration + (1 | Subject) + (1 | Instance)
	SR[e] ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)
Diagram	FD ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)
	FC ~ Time + Size + Duration + (1 | Subject) + (1 | Instance)
	SR ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)
Instructor and diagram	saccade[f] ~ Time + (1|subject)

Note. [a]Log-transformed total fixation duration within each instance.
[b]Log-transformed AOI instance size. [c]Log-transformed AOI instance duration.
[d]Fixation count within each instance. [e]Instance skip rate.
[f]Number of integrative saccades between the instructor and diagram AOIs.

Table 26 presents the results of the linear mixed-effects models. It illustrates that the time spent on the instructor AOI significantly increased between times 1 and 3. No significant difference was found between times 1 and 2 or between times 2 and 3. In contrast, the time spent on the diagram AOI declined from time 1 to times 2 and 3. There was no significant difference between times 2 and 3.

Table 26 Results of Within-Group Comparisons for the Linear Mixed-Effects Models for the Instructor and the Diagram AOIs

Variable	*b*	95% CI	*SE*	*t*	*p*	R^2m	R^2c
Total fixation duration on the instructor AOI							
Time 1 – Time 2	0.05	[-0.01, 0.11]	0.03	1.55	.14	.35	.42
Time 2 – Time 3	0.03	[-0.03, 0.09]	0.03	1.01	.32	.35	.42
Time 1 – Time 3	0.07	[0.01, 0.14]	0.03	2.18	**.04**	.35	.42
Total fixation duration on the diagram AOI							
Time 1 – Time 2	-0.06	[-0.11, 0.00]	0.03	-2.13	**.04**	.36	.38
Time 2 – Time 3	0.01	[-0.03, 0.05]	0.02	0.46	.65	.36	.38
Time 1 – Time 3	-0.05	[-0.09, 0.01]	0.02	-2.31	**.03**	.36	.38

Note. Values of *p* in boldface are significant at the .05 level of alpha.

As summarized in Table 27, the results of the Poisson mixed-effects models indicated that time was not a significant predictor for the repetition group's fixation count to the instructor AOI. Their fixation count to the diagram AOI, however, significantly declined from times 2 to 3 and from times 1 to 3. The results of logistic mixed-effects models suggested that the odds of skipping the instances of the instructor AOI significantly increased between times 1 and 2. The odds of skipping the instances of the diagram AOI also significantly rose from time 1 to times 2 and 3. In terms of covariates, the models for the instructor AOI found that greater AOI instance size was linked to decreased fixation durations and counts but increased instance skip rates. Interestingly, the models constructed for the diagram AOI revealed an opposite trend, showing that greater AOI instance size was related to more fixation counts but reduced instance skip rates. The results also found that longer AOI instance duration was associated with more fixation durations and counts but lower instance skip rates of the instructor and the diagram AOIs. Lastly, the Poisson mixed-effects model analyzing the number of integrative saccades between the instructor and the diagram AOIs revealed a main effect of time: the number of integrative saccades declined between times 1 to 2 but increased from times 2 to 3. No significant difference was found between times 1 and 3. For full model summaries, see Appendix 12.

Table 27 Results of Within-Group Comparisons for the Generalized Linear Mixed-Effects Models for the Instructor and the Diagram AOIs

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Fixation count on the instructor AOI									
Time 1 – Time 2	0.01	[-0.04, 0.06]	0.03	1.01	[0.96, 1.06]	0.35	.73	.09	.10
Time 2 – Time 3	0.03	[-0.02, 0.08]	0.02	1.03	[0.98, 1.08]	1.27	.20	.09	.10
Time 1 – Time 3	0.04	[-0.01, 0.09]	0.02	1.04	[0.99, 1.09]	1.70	.09	.09	.10
Fixation count on the diagram AOI									
Time 1 – Time 2	-0.002	[-0.03, 0.03]	0.02	1.00	[0.97, 1.03]	-0.12	.90	.10	.11
Time 2 – Time 3	-0.03	[-0.06, 0.00]	0.02	0.97	[0.93, 1.00]	-2.08	**.04**	.10	.11
Time 1 – Time 3	-0.04	[-0.07, -0.01]	0.02	0.96	[0.93, 0.99]	-2.32	**.02**	.10	.11
Instance skip rate on the instructor AOI									
Time 1 – Time 2	0.29	[0.04, 0.54]	0.13	1.34	[1.04, 1.72]	2.28	**.02**	.08	.38
Time 2 – Time 3	-0.25	[-0.58, 0.08]	0.17	0.78	[0.56, 1.08]	-1.50	.13	.08	.38
Time 1 – Time 3	0.04	[-0.42, 0.51]	0.24	1.04	[0.66, 1.66]	0.18	.86	.08	.38
Instance skip rate on the diagram AOI									
Time 1 – Time 2	0.53	[0.24, 0.82]	0.15	1.70	[1.26, 2.27]	3.53	**< .001**	.03	.26
Time 2 – Time 3	-0.01	[-0.30, 0.27]	0.15	0.99	[0.74, 1.31]	-0.09	.93	.03	.26
Time 1 – Time 3	0.52	[0.04, 0.99]	0.24	1.68	[1.04, 2.69]	2.14	**.03**	.03	.26
Number of integrative saccades between the instructor and the diagram AOIs									
Time 1 – Time 2	-0.24	[-0.30, -0.18]	0.03	0.79	[0.74, 0.84]	-7.62	**<.001**	.09	.89
Time 2 – Time 3	0.23	[0.17, 0.29]	0.03	1.26	[1.19, 1.34]	7.34	**<.001**	.09	.89
Time 1 – Time 3	-0.01	[-0.07, 0.05]	0.03	0.99	[0.94, 1.05]	-0.28	.78	.09	.89

Note. Values of *p* in boldface are significant at the .05 level of alpha.

4.3 RQ2: Cognitive Processes Engaged in Repeated Viewing

In total, 15 participants made 544 stimulated comments, which were segmented into 657 chunks based on the coding scheme for analyzing participants' cognitive processes and strategy use. Table 28 presents the number of chunks counted for each category, namely, listening comprehension, reading comprehension, visual comprehension, integration of information, listening/viewing strategies, note-viewing, and processing of social cues. In addition to the raw counts, percentages of each category were also computed by dividing the number of codes in each category by the total number of codes. Percentages of subcodes of listening comprehension and listening/viewing strategies were presented as well, calculated as the number of subcodes divided by the total number of codes in their respective categories. Participants who performed the task once (group 1) and twice (group 2) made more comments than those who performed the task three times (group 3). A large part of comments related to listening comprehension, with group 1 reporting listening processes more frequently than groups 2 and 3. While the three groups reported a similar proportion of lower-level listening processes, group 1 made more comments about higher-level listening processes. Regarding reading comprehension, group 2 mentioned reading diagram labels slightly more often than the other two groups. In terms of visual comprehension, there was no notable difference in either lower- or higher-level processes across the three groups. Group 1 also reported engaging in more cognitive processes relating to the integration of information from different modalities than the other two groups.

Another major part of the comments related to strategy use. Overall, groups 2 and 3 relied notably more on listening/viewing strategies than group 1. To be more specific, group 3 mentioned using cognitive strategies slightly more frequently than groups 1 and 2; Group 2 described more behaviors relating to using metacognitive strategies, in particular, monitoring comprehension during task performance; the socio-affective strategy (i.e., managing one's emotions) was only reported by a participant in group 3; and group 2 made particularly more comments about using task-specific strategies, such as selecting important information, remembering target word pronunciation and the instructor's exact words. In addition, both groups 2 and 3 reported reading their notes during repeated viewing. As compared to groups 1 and 2, group 3 more frequently mentioned paying conscious attention to the instructor's nonverbal communication cues.

Table 28 Code Frequency for Stimulated Recall Comments Regarding Cognitive Processes and Strategy Use by Group

Code	Group 1 (n = 5)	Group 2 (n = 5)	Group 3 (n = 5)
Listening comprehension	144 (62.88%)	97 (40.76%)	88 (46.32%)
Lower-level processes	53 (23.14%)	45 (18.91%)	41 (21.58%)
Not specified	14	8	4
Lexical search	25	26	22
Parsing	14	11	15
Higher-level processes	91 (39.74%)	52 (21.85%)	47 (24.74%)
Meaning construction	53	20	26
Discourse construction	38	32	21
Reading comprehension Graphemic input analysis (lower level)	15 (6.55%)	23 (9.66%)	11 (5.79%)
Visual comprehension	12 (5.24%)	11 (4.62%)	7 (3.68%)
Visual feature analysis (lower level)	4	6	2
Depictive processing (higher level)	8	5	5
Integration of information	34 (14.85%)	15 (6.30%)	24 (12.04%)
Perceptual processing (lower-level)	18	6	13
Semantic processing (higher level)	16	9	11
Listening/Viewing strategies	20 (8.73%)	82 (34. 45%)	39 (20.94%)
Cognitive strategies	17 (7.42%)	23 (9.66%)	22 (12.63%)
Avoidance	5	5	6

Code	Group 1 (n = 5)	Group 2 (n = 5)	Group 3 (n = 5)
Linguistic inferencing	5	2	4
Kinesic inferencing	1	8	1
Elaboration	0	1	6
Prediction	4	0	0
Contextualizing	0	0	1
Reorganizing notes	0	7	4
Translating	2	0	0
Metacognitive strategies	1 (0.44%)	35 (14.71%)	12 (6.32%)
Planning	0	4	4
Focusing attention	1	9	2
Monitoring	0	22	5
Evaluation	0	0	1
Socio-affective strategy	0 (0%)	0 (0%)	2 (1.05%)
Task-specific strategies	2 (0.88%)	24 (10.08%)	3 (1.58%)
Selecting information	1	13	3
Remembering pronunciation	1	4	0
Remembering the instructor's exact words	0	7	0
Note viewing	0 (0%)	5 (2.10%)	6 (3.16%)
Processing of social cues	4 (1.75%)	5 (2.10%)	15 (7.89%)
Total	229	238	190

4.4 RQ3: Relationship between Processing and Comprehension

In order to explore how task repetition affected the relationship between learners' viewing behavior and lecture comprehension, linear mixed-effects models were fitted, with total fixation duration on the instructor and the diagram AOIs as the fixed effects, and subject and AOI instance as random effects. Accuracy percentages were computed for the number of correct and incorrect idea units recalled for each video (i.e., dividing the number of correct/incorrect idea units by the total number of idea units in the video transcript and multiplying the quotient by 100), given that the three video transcripts contained different numbers of idea units. As previously stated, I only focused on the eye-movement data extracted from the first video in the current study, therefore, the dependent variables were percentages of correct and incorrect idea units recalled for the first video (see Table 29 for descriptive statistics). Two model structures were first fitted to the data: the null model only included random intercepts for subjects and AOI instances, and the second model contains the control group's total fixation duration on the instructor AOI as the predictor.

Null model: correctIU% ~ (1 | Subject) + (1 | Instance)

Model 1: correctIU% ~ InstructorFD + (1 | Subject) + (1 | Instance)

When fitting the models, an error message was generated, suggesting that the variance due to random effects was too small. The model summaries also showed that the variance of the random intercepts for subjects as well as instances was close to zero. Consequently, the random effects were dropped due to the lack of subject- and instance-level variance, and linear regression models were constructed instead. Participants' CAE scores were added as a participant-level covariate in the analysis but were removed because they did not significantly improve the model fit. As shown in Table 30, the control group's total fixation duration on the instructor was a significant predictor for the number of correct idea units, with shorter fixation durations on the instructor being linked to more correctly recalled idea units. No significant difference was found between their total fixation duration on the diagram and the number of correct idea units recalled. The control group's total fixation duration on the instructor or the diagram AOI did not emerge as a significant predictor of the number of incorrect idea units.

Table 29 Descriptive Statistics for Accuracy Percentages on the Number of Idea Units Recalled for the First Video by Group

Group	Percentage of correct idea units		Percentage of incorrect idea units	
	M (SD)	95% CI	*M (SD)*	95% CI
Control group (n = 30)	23.56% (12.36)	[18.94, 28.17]	4.13% (3.04)	[3.00, 5.27]
Repetition group (n = 30)	39.69% (13.05)	[34.81, 44.56]	2.44% (2.55)	[1.49, 3.40]

Table 30 Results of the Linear Regression Models Examining the Relationship between the Control Group's Viewing Behavior and Lecture Comprehension

Variable	*b*	95% CI	*SE*	*t*	*p*	R^2
Percentage of correct idea units –Total fixation duration on the instructor AOI	-11.03	[-17.06, -5.00]	2.94	-3.75	**.001**	.33
Percentage of correct idea units – Total fixation duration on the diagram AOI	-13.64	[-30.13, 2.85]	8.05	-1.69	.10	.09
Percentage of incorrect idea units – Total fixation duration on the instructor AOI	-0.29	[-2.10, 1.53]	0.89	-0.32	.75	<.01
Percentage of incorrect idea units – Total fixation duration on the diagram AOI	-0.97	[-5.21, 3.27]	2.07	-0.47	.64	.01

Note. Values of *p* in boldface are significant at the .05 level of alpha.

Linear mixed-effects models were also constructed with the repetition group's total fixation duration on the instructor and the diagram AOIs at each exposure and within the whole treatment as the fixed effect, respectively. Subject and AOI instance were added to the models as random intercepts, and the dependent variables were again the percentage of correct and incorrect idea units recalled for the first video. The same warning message was displayed after fitting the mixed-effects models, and so random effects were removed. The results of linear regression are presented in Table 31, showing that the less time the repetition group spent looking at the instructor at times 2 and 3, the more correct idea units they recalled. A weak negative relationship between the cumulative fixation duration on the instructor AOI and the percentage of correct idea units was also found, but it did not reach a significant level. The results presented in Table 32 indicated that the number of correct idea units was significantly predicted by the repetition group's total fixation duration on the diagram AOI at time 3, with less time spent looking at the diagram leading to more correct idea units. None of the fixation duration indices, however, predicted the percentage of incorrect idea units. For full model summaries, see Appendix 13.

Table 31 Results of the Linear Regression Models Examining the Relationship between the Repetition Group's Fixation Duration on the Instructor and Lecture Comprehension

Variable	*b*	95% CI	*SE*	*t*	*p*	R^2
Percentage of correct idea units – Time 1 total fixation duration	1.54	[-6.46, 9.54]	3.91	0.39	.70	.01
Percentage of correct idea units – Time 2 total fixation duration	-6.05	[-11.61, -0.48]	2.72	-2.23	**.03**	.15
Percentage of correct idea units – Time 3 total fixation duration	-3.01	[-5.14, -0.87]	1.04	-2.89	**.01**	.22
Percentage of correct idea units – Cumulative fixation duration	-7.24	[-14.91, 0.43]	3.75	-1.93	.06	.11
Percentage of incorrect idea units – Time 1 total fixation duration	-0.33	[-1.90, 1.23]	0.76	-0.44	.67	.01
Percentage of incorrect idea units – Time 2 total fixation duration	-0.25	[-1.43, 0.92]	0.57	-0.44	.66	.01
Percentage of incorrect idea units – Time 3 total fixation duration	0.06	[-0.41, 0.54]	0.23	0.27	.79	<.01
Percentage of incorrect idea units – Cumulative fixation duration	-0.22	[-1.81, 1.38]	0.78	-0.28	.78	<.01

Note. Values of *p* in boldface are significant at the .05 level of alpha.

Table 32 Results of the Linear Regression Models Examining the Relationship between the Repetition Group's Fixation Duration on the Diagram and Lecture Comprehension

Variable	*b*	95% CI	*SE*	*t*	*p*	R^2
Percentage of correct idea units – Time 1 total fixation duration	-1.05	[-15.58, 13.48]	7.09	-0.15	.88	<.01
Percentage of correct idea units – Time 2 total fixation duration	-7.44	[-16.93, 2.04]	4.63	-1.61	.12	.08
Percentage of correct idea units – Time 3 total fixation duration	-4.54	[-8.79, -0.29]	2.07	-2.19	**.04**	.14
Percentage of correct idea units – Cumulative fixation duration	-11.32	[-25.74, 3.10]	7.04	-1.61	.12	.08
Percentage of incorrect idea units – Time 1 total fixation duration	-0.21	[-3.05, 2.63]	1.38	-0.15	.88	<.01
Percentage of incorrect idea units – Time 2 total fixation duration	0.73	[-1.18, 2.65]	0.94	0.79	.44	.02
Percentage of incorrect idea units – Time 3 total fixation duration	0.13	[-0.76, 1.03]	0.44	0.31	.76	<.01
Percentage of incorrect idea units – Cumulative fixation duration	0.65	[-2.28, 3.59]	1.43	0.46	.65	.01

Note. Values of *p* in boldface are significant at the .05 level of alpha.

4.5 RQ4: Vocabulary Acquisition

To address RQ4 (To what extent does repeating a multimodal lecture-viewing task affect learners' incidental acquisition of technical words, as measured by offline vocabulary tests?), participants' vocabulary post-test scores under different conditions (viewing once and viewing three times) were explored. Six logistic mixed-effects models were constructed with group (control and repetition group) served as the fixed effect (see Table 33 for best-fit model structures). Subject and item were added to the models as random intercepts, and by-item random slopes for group were included in some models. Participant-level (i.e., CAE scores) and item-level covariates (i.e., POS) were also included as they significantly improved model fit.

Table 33 Model Structures for the Target Word AOIs

Dependent variable	Model structure
Immediate form recognition scores	FReco1 ~ Group + (1 | Subject) + (1 + Group | Item)
Immediate meaning recognition	MReco1 ~ Group + CAE[a] + POS[b] + (1 | Subject) + (1 | Item)
Immediate meaning recall	MReca1 ~ Group + (1 | Subject) + (1 + Group | Item)
Delayed from recognition	FReco2 ~ Group + (1 | Subject) + (1 | Item)
Delayed meaning recognition	MReco2 ~ Group + CAE + POS + (1 | Subject) + (1 | Item)
Delayed meaning recall	MReca2 ~ Group + CAE + (1 | Subject) + (1 + Group | Item)

Note. [a]Log-transformed Cambridge Certificate in Advanced English test scores.
[b]Part of speech.

Descriptive statistics for the immediate and delayed vocabulary post-test scores are presented in Table 34. It can be observed that the repetition group scored higher than the control group on all tests. As shown in Table 35, the results of the logistic mixed-effects models confirmed a significant main effect of group on participants' immediate and delayed vocabulary gains. The CAE test scores and the POS of target words, included as covariates, emerged as significant predictors for the vocabulary post-test scores. Specifically, higher CAE scores related to greater immediate and delayed meaning recognition, as well as delayed meaning recall. Adjectives were linked to increased immediate and delayed meaning recognition, whereas nouns were linked to decreased meaning recognition. A summary of the logistic mixed-effects models can be found in Appendix 14.

Table 34 Descriptive Statistics for the Immediate and Delayed Vocabulary Post-Tests by Group

Group	Immediate post-test						Delayed post-test					
	Form recognition		Meaning recognition		Meaning recall		Form recognition		Meaning recognition		Meaning recall	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Control group (n = 30)	16.3 (2.25)	[15.46, 17.14]	6.3 (2.23)	[5.47, 7.13]	1.91 (1.51)	[1.34, 2.47]	16.67 (2.45)	[15.75, 17.58]	6.17 (2.23)	[5.33, 7.00]	0.99 (1.51)	[0.43, 1.56]
Repetition group (n = 30)	18.7 (2.28)	[17.85, 19.55]	8.83 (1.82)	[8.15, 9.51]	4.44 (2.23)	[3.60, 5.27]	18.77 (1.87)	[18.07, 19.46]	7.87 (2.26)	[7.02, 8.71]	2.45 (1.78)	[1.79, 3.12]

Table 35 Results of Between-Group Comparisons for the Immediate and Delayed Vocabulary Post-Tests

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Immediate vocabulary post-test									
Form recognition	0.53	[0.07, 1.00]	0.24	1.70	[1.07, 2.69]	2.24	**.03**	.02	.24
Meaning recognition	1.36	[0.83, 1.90]	0.27	3.89	[2.28, 6.62]	5.00	< **.001**	.18	.31
Meaning recall	1.45	[0.88, 2.03]	0.29	4.28	[2.40, 7.61]	4.94	< **.001**	.12	.33
Delayed vocabulary post-test									
Form recognition	0.71	[0.29, 1.12]	0.21	2.00	[1.31, 3.06]	3.34	**.001**	.02	.37
Meaning recognition	0.86	[0.34, 1.39]	0.27	2.37	[1.40, 4.03]	3.21	**.001**	.13	.29
Meaning recall	1.46	[0.75, 2.18]	0.36	4.25	[1.97, 9.17]	4.01	< **.001**	.15	.37

Note. Values of *p* in boldface are significant at the .05 level of alpha.

4.6 RQ5: Processing of Vocabulary during Repeated Viewing

In response to RQ5 (To what extent does repeating a multimodal lecture-viewing task affect learners' visual attention to technical words, as reflected in their eye movements?), the repetition group's eye movements to the target word AOIs across three exposures were analyzed. Two linear mixed-effect models were constructed for continuous dependent variables (i.e., total fixation duration and mean fixation duration). Two Poisson mixed-effects models were built for count-dependent variables (i.e., fixation count and run count), and a logistic mixed-effects model was built for AOI skip rate (i.e., whether an AOI was fixed upon: an AOI was considered skipped if no fixation was fixated upon). All final models had by-subject and by-random intercepts, and some models included by-subject random slopes for time. Participants' VST scores, AOI size (i.e., the weighted mean size), the number of nonverbal signals, and the length of target items were added to the model as covariates. See Table 36 for the best-fit model structures.

Table 36 Model Structures for the Target Word AOIs

Dependent variable	Model structure
Log-transformed total fixation duration	FD ~ Time + Size[a] + Nonverbal[b] + (1 + Time | Subject) + (1 | Item)
Log-transformed mean fixation duration	MFD ~ Time + Length[c] + (1 | Subject) + (1 | Item)
Fixation count	FC ~ Time + Size + (1 | obs_effect) + (1 + Time | Subject) + (1 | Item)
Run count	RC ~ Time + Size + (1 | obs_effect) + (1 + Time | Subject) + (1 | Item)
AOI Skip rate	SR ~ Time + VST[d] + Size + Nonverbal + (1 | Subject) + (1 | Item)

Note. [a]Log-transformed target word AOI size. [b]Number of nonverbal signals. [c]Length of Target words. [d]Log-transfored VST scores.

Table 37 presents the descriptive statistics for the repetition group's processing of the target words over repeated viewing. As demonstrated in Table 38, the results of the linear mixed-effects models revealed that the time spent on the target words, as measured by total fixation duration and mean fixation duration, significantly decreased from time 1 to times 2 and 3, but no significant difference was found between times 2 and 3.

The results of the Poisson mixed-effects models are presented in Table 39. The results revealed that the fixation count and the run count were significantly higher at time 1 than at times 2 and 3, and no significant difference was found between times 2 and 3. Table 39 also shows the results of the logistic mixed-effects model, indicating that the odds of skipping the target words significantly increased between times 1 and 2, times 1 and 3, and times 2 and 3. Regarding covariates, the six models found that greater AOI size was related to longer fixation durations, higher fixation and run counts, and lower skip rate; more non-verbal signals were linked to increased fixation durations and decreased skip rate; greater word length was associated with longer mean fixation duration; and participants with higher vocabulary size skipped more target words. A summary of the models can be found in Appendix 15.

Table 37 Descriptive Statistics for the Target Word Eye-movement Measures by Time

Variable	Time 1		Time2		Time 3	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Total fixation duration (milliseconds)	4856 (4152)	[4406, 5305]	3354 (3428)	[2983, 3725]	3017 (3619)	[2525, 3409]
Mean fixation duration (milliseconds)	308.16 (141.16)	[292.87, 323.44]	283.71 (176.52)	[264.59, 302.83]	247.93 (148.61)	[231.84, 264.03]
Fixation count	173.33 (45.64)	[156.29, 190.38]	124.80 (42.23)	[109.03, 140.57]	112.83 (49.54)	[94.33, 131.33]
Run count	112.23 (27.79)	[101.86, 122.61]	86.67 (28.39)	[76.07, 97.27]	77.90 (32.33)	[65.83, 78.31]
AOI skip rate	0.02 (0.12)	[0.00, 0.03]	0.05 (0.21)	[0.02, 0.07]	0.14 (0.35)	[0.10, 0.18]

Table 38 Results of Within-Group Comparisons for the Linear Mixed-Effects Models for the Target Word AOIs

Variable	*b*	95% CI	*SE*	*t*	*p*	R^2m	R^2c
Total fixation duration							
Time 1 – Time 2	-0. 49	[-0.66, -0.32]	0.09	-5.51	< **.001**	.42	.63
Time 2 – Time 3	-0.10	[-0.28, 0.07]	0.09	-1.19	.24	.42	.63
Time 1 – Time 3	-0.59	[-0.83, -0.36]	0.12	-4.98	< **.001**	.42	.63
Mean fixation duration							
Time 1 – Time 2	-0.07	[-0.12, -0.02]	0.03	-2.56	**.011**	.03	.17
Time 2 – Time 3	-0.02	[-0.07, 0.04]	0.03	-0.63	.53	.03	.17
Time 1 – Time 3	-0.09	[-0.14, -0.03]	0.03	-3.12	**.002**	.03	.17

Note. Values of *p* in boldface are significant at the .05 level of alpha.

Table 39 Results of Within-Group Comparisons for the Generalized Mixed-Effects Models for the Target Word AOIs

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Fixation count									
Time 1 – Time 2	-0.38	[-0.50, -0.26]	0.06	0.69	[0.61, 0.77]	-6.24	< **.001**	.43	.92
Time 2 – Time 3	-0.09	[-0.21, 0.03]	0.06	0.91	[0.81, 1.03]	-1.54	.12	.43	.92
Time 1 – Time 3	-0.47	[-0.65, -0.30]	0.09	0.62	[0.52, 0.74]	-5.23	< **.001**	.43	.92
Run count									
Time 1 – Time 2	-0.29	[-0.40, -0.18]	0.06	0.75	[0.67, 0.84]	-5.08	< **.001**	.38	.84
Time 2 – Time 3	-0.09	[-0.20, -0.01]	0.05	0.91	[0.82, 1.01]	-1.71	.09	.38	.84
Time 1 – Time 3	-0.38	[-0.53, -0.22]	0.08	0.69	[0.59, 0.80]	-4.81	< **.001**	.38	.84
AOI skip rate									
Time 1 – Time 2	1.49	[0.33, 2.64]	0.59	4.44	[1.39,14.14]	2.52	**.01**	.52	.63
Time 2 – Time 3	1.52	[0.84, 2.20]	0.35	4.56	[2.31, 9.01]	4.38	< **.001**	.52	.63
Time 1 – Time 3	3.00	[1.90, 4.10]	0.56	20.24	[6.74, 60.80]	5.35	< **.001**	.52	.63

Note. Values of *p* in boldface are significant at the .05 level of alpha.

4.7 RQ6: Awareness and Vocabulary Learning

Among 544 comments made in total, 124 comments concerned the target words and were segmented into 210 chunks based on the coding scheme for analyzing participants' awareness of the target words. Table 40 shows the number of chunks counted for each level of awareness, namely, noticing and understanding. Apart from the raw counts, percentages of each category were also calculated by dividing the number of codes in each category by the total number of codes. The majority of the comments related to understanding a target word across the three groups, but groups 1 and 3 reported noticing a target word more frequently than group 2. Despite the small sample size, all groups reported noticing specific form-related features, with groups 2 and 3 reporting more noticing of orthographic forms than group 1. The chunks were also coded for sources of awareness (see Table 41 for results). Percentage of subcodes (e.g., "not specified", "from aural commentaries", "from diagram labels") in each category (i.e., noticing and understanding) was computed as the number of subcodes divided by the total number of codes. Across all three groups, participants mentioned diagram labels as a major source of noticing the target words, while aural commentaries were specified as the main source for understanding target word meanings. As compared to group 1, groups 2 and 3 made notably more comments about relying on integrated information to understand target word meanings. Notes were reported as an important source of input by these two groups.

Table 40 Code Frequency for Stimulated Recall Comments about Level of Awareness by Group

Code	Group 1 (n = 5)	Group 2 (n = 5)	Group 3 (n = 5)
Noticing	32 (40%)	31 (36.90%)	24 (42.86%)
Not specified	27	20	14
Orthographic form	2	9	8
Phonological form	2	2	0
GPC	1	0	0
POS	0	0	2
Understanding	48 (60%)	53 (63.10%)	32 (57.14%)
Total	80	84	56

Table 41 Code Frequency for Stimulated Recall Comments about Input Sources by Group

Code	Group 1 (n = 5)	Group 2 (n = 5)	Group 3 (n = 5)
Noticing			
Not specified	14 (17.50%)	3 (3.57%)	5 (8.93%)
From aural commentaries	4 (5.00%)	7 (8.33%)	4 (7.14%)
From diagram labels	13 (16.25%)	11 (13.10%)	12 (21.43%)
From multiple input sources	1 (1.25%)	0 (0%)	0 (0%)
From notes	0 (0%)	10 (11.90%)	3 (5.36%)
Understanding			
Not specified	2 (2.50%)	7 (8.33%)	3 (5.36%)
From aural commentaries	41 (51.25%)	35 (41.67%)	21 (37.5%)
From diagram illustrations	3 (3.75%)	0 (0%)	0 (0%)
From multiple input sources	2 (2.50%)	10 (11.90%)	6 (10.71%)
From notes	0 (0%)	0 (0%)	1 (1.79%)
From nonverbal signals	0 (0%)	1 (1.19%)	1 (1.79%)
Total	80	84	56

4.8 RQ7: Relationship between Processing and Learning

Finally, in order to examine the effect of task repetition on the relationship between participants' visual attention to the target items and vocabulary learning gains, a set of logistic mixed-effects models was constructed (see Table 42 for model structures). The fixed effect included the control group's total fixation duration on the target words, and outcome variables were their vocabulary post-test (i.e., immediate and delayed form recognition, meaning recognition, and meaning recall) scores. Another set of logistic mixed-effects models was built, with the repetition group's total fixation duration at three exposure and during the entire treatment serving as predictors of their post-test results (see Table 43 for model structures). The participants' CAE scores, POS of the target words, number of nonverbal signals, and target word AOI size were entered into some of the models as covariates.

Table 42 Structures of the Logistic Mixed-Effects Models Examining the Relationship between the Control Group's Visual Attention to the Target Words and Vocabulary Gains

Dependent variable	Model structure
Immediate form recognition	FReco1 ~ FD[a] + (1 | Subject) + (1 | Item)
Immediate meaning recognition	MReco1 ~ FD + POS[b] + (1 | Subject) + (1 | Item)
Immediate meaning recall	MReca1 ~ FD + (1 | Subject) + (1 | Item)
Delayed form recognition	FReco2 ~ FD + Nonverbal[c] + (1 | Subject) + (1 | Item)
Delayed meaning recognition	MReco2 ~ FD + POS + (1 | Subject) + (1 | Item)
Delayed meaning recall	MReca2 ~ FD + CAE[d] + POS + (1 | Subject) + (1 | Item)

Note. [a]Log-transformed total fixation duration on the target word AOIs. [b]Part of speech. [c]Number of nonverbal signals. [d]Cambridge Certificate in Advanced English test scores.

Table 43 Structures of the Logistic Mixed-Effects Models Examining the Relationship between the Repetition Group's Visual Attention to the Target Words and Vocabulary Gains

Dependent variable	Model structure
Immediate form recognition	FReco1 ~ FD1[a] + (1 | Subject) + (1 | Item)
	FReco1 ~ FD2[b] + (1 | Subject) + (1 | Item)
	FReco1 ~ FD3[c] + (1 | Subject) + (1 | Item)
	FReco1 ~ CFD[d] + (1 | Subject) + (1 | Item)
Immediate meaning recognition	MReco1 ~ FD1 + CAE[e] + Nonverbal[f] + (1 | Subject) + (1 | Item)
	MReco1 ~ FD2 + Nonverbal + (1 | Subject) + (1 | Item)
	MReco1 ~ FD3 + POS[g] + (1 | Subject) + (1 + | Item)
	MReco1 ~ CFD + CAE + Nonverbal + (1 | Subject) + (1 | Item)
Immediate meaning recall	MReca1 ~ FD1 + AOI[h] + (1 + FD1 | Subject) + (1 | Item)
	MReca1 ~ FD2 + CAE + (1 | Subject) + (1 | Item)
	MReca1 ~ FD3 + (1 | Subject) + (1 | Item)
	MReca1 ~ CFD + (1 + CFD | Subject) + (1 | Item)
Delayed form recognition	FReco2 ~ FD1 + (1 | Subject)
	FReco2 ~ FD2 + (1 | Subject)
	FReco2 ~ FD3 + (1 | Subject)
	FReco2 ~ CFD + (1 | Subject)
Delayed meaning recognition	MReco2 ~ FD1 + Nonverbal + (1 | Subject) + (1 | Item)
	MReco2 ~ FD2 + Nonverbal + (1 | Subject) + (1 | Item)
	MReco2 ~ FD3 + Nonverbal + (1 | Subject) + (1 | Item)
	MReco2 ~ CFD + Nonverbal + (1 | Subject) + (1 | Item)
Delayed meaning recall	MReca2 ~ FD1 + (1 | Subject) + (1 | Item)
	MReca2 ~ FD2 + (1 | Subject)
	MReca2 ~ FD3 + (1 | Subject) + (1 | Item)
	MReca2 ~ CFD + (1 + CFD | Subject) + (1 | Item)

Note. [a]Log-transformed total fixation duration on the target word AOIs at time 1.

[b]Log-transformed total fixation duration on the target word AOIs at time 2.

[c]Log-transformed total fixation duration on the target word AOIs at time 3.

[b]Log-transformed cumulative fixation duration on the target word AOIs during the whole treatment.

[e]Log-transformed Cambridge Certificate in Advanced English test scores.

[f]Number of nonverbal signals. [g]Part of speech. [h]Log-transformed AOI size.

Tables 44, 45, and 46 show that neither of the fixation duration indices (i.e., the control group's total fixation duration on the target words and the repetition group's total fixation duration on the target words at times 1, 2, 3, and the cumulative fixation duration during the whole treatment, respectively) were significant predictors of the control or repetition group's vocabulary gains. The covariates, including the CAE test scores, the POS of the target words, and the number of nonverbal signals, emerged as predictors of the post-test scores. In particular, higher CAE scores were linked to greater delayed meaning recall for the control group, immediate meaning recognition for the repetition group during the first viewing and the entire treatment, as well as immediate meaning recall for the repetition group during the second viewing. Adjectives were linked to increased immediate and delayed meaning recognition and delayed meaning recall for the control group, as well as increased immediate meaning recognition for the repetition group during the third viewing. In addition, the number of nonverbal signals was associated with increased delayed form recognition for the control group and increased immediate meaning recognition for the repetition group during the first and second viewing as well as during the whole treatment. The number of nonverbal signals was also linked to increased delayed meaning recognition for the repetition group at times 1, 2, 3, and during the whole treatment. Lastly, smaller target word AOIs were found to be related to greater immediate meaning recall for the repetition group during the first viewing. For model summaries, see Appendix 16.

Table 44 Relationship Between the Control Group's Processing of the Target Words and Vocabulary Post-Test Scores

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Immediate post-test									
Form recognition – Total fixation duration	0.19	[-0.10, 0.48]	0.15	1.21	[0.90, 1.61]	1.26	.21	.01	.14
Meaning recognition – Total fixation duration	0.09	[-0.16, 0.32]	0.12	1.09	[0.86, 1.38]	0.71	.48	.07	.17
Meaning recall – Total fixation duration	0.10	[-0.13, 0.33]	0.12	1.10	[0.88, 1.38]	0.83	.40	.00	.26
Delayed post-test									
Form recognition – Total fixation duration	-0.29	[-0.83, 0.24]	0.27	0.75	[0.44, 1.27]	-1.07	.28	.36	.70
Meaning recognition – Total fixation duration	0.13	[-0.13, 0.39]	0.13	1.14	[0.88, 1.48]	1.01	.31	.05	.19
Meaning recall – Total fixation duration	0.19	[-0.11, 0.53]	0.16	1.21	[0.88, 1.66]	1.20	.23	.14	.38

Table 45 Relationship between the Repetition Group's Processing of the Target Words and Vocabulary Immediate Post-Test Scores

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Form recognition – Time 1 fixation duration	0.25	[-0.07, 0.58]	0.17	1.29	[0.93, 1.78]	1.52	.13	.02	.08
Form recognition – Time 2 fixation duration	0.09	[-0.21, 0.40]	0.15	1.10	[0.81, 1.48]	0.59	.55	<.01	.02
Form recognition – Time 3 fixation duration	0.17	[-0.14, 0.49]	0.16	1.19	[0.87, 1.64]	1.08	.28	.01	0.09
Form recognition – Cumulative fixation duration	0.25	[-0.13, 0.64]	0.20	1.29	[0.88, 1.89]	1.29	.20	.02	.09
Meaning recognition – Time 1 fixation duration	0.03	[-0.33, 0.40]	0.19	1.03	[0.72, 1.49]	0.18	.86	.16	.32
Meaning recognition – Time 2 fixation duration	-0.08	[-0.40, 0.25]	0.17	0.93	[0.67, 1.29]	-0.45	.65	.12	.30
Meaning recognition – Time 3 fixation duration	-0.11	[-0.44, 0.22]	0.17	0.90	[0.64, 1.24]	-0.66	.51	.07	.23
Meaning recognition – Cumulative fixation duration	0.04	[-0.39, 0.51]	0.23	1.04	[0.67, 1.61]	0.17	.87	.16	.32
Meaning recall – Time 1 fixation duration	0.19	[-0.05, 0.42]	0.12	1.21	[0.95, 1.53]	1.55	.12	.01	.24
Meaning recall – Time 2 fixation duration	-0.01	[-0.18, 0.16]	0.09	0.99	[0.84, 1.17]	-0.13	.90	.03	.21
Meaning recall – Time 3 fixation duration	-0.04	[-0.20, 0.12]	0.08	0.96	[0.82, 1.13]	-0.45	.65	<.001	.20
Meaning recall – Cumulative fixation duration	0.03	[-0.13, 0.33]	0.15	1.03	[0.77, 1.38]	0.19	.85	<.01	.26

Table 46 Relationship between the Repetition Group's Processing of the Target Words and Vocabulary Delayed Post-Test Scores

Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	R^2m	R^2c
Form recognition – Time 1 fixation duration	0.21	[-0.24, 0.67]	0.23	1.24	[0.78, 1.96]	0.91	.36	<.01	.06
Form recognition – Time 2 fixation duration	0.08	[-0.39, 0.55]	0.24	1.08	[0.68, 1.74]	0.34	.74	<.01	.18
Form recognition – Time 3 fixation duration	0.32	[-0.17, 0.81]	0.25	1.37	[0.84, 2.24]	1.27	.20	.03	.29
Form recognition – Cumulative fixation duration	0.40	[-0.08, 0.88]	0.25	1.49	[0.92, 2.41]	1.63	.10	.03	.28
Meaning recognition – Time 1 fixation duration	-0.02	[-0.37, 0.33]	0.18	0.98	[0.69, 1.39]	-0.10	.92	.18	.40
Meaning recognition – Time 2 fixation duration	-0.21	[-0.54, 0.12]	0.17	0.81	[0.58, 1.13]	-1.25	.21	.19	.42
Meaning recognition – Time 3 fixation duration	-0.07	[-0.44, 0.29]	0.19	0.93	[0.65, 1.34]	-0.40	.69	.20	.49
Meaning recognition – Cumulative fixation duration	-0.31	[-0.74, 0.10]	0.21	0.73	[0.49, 1.10]	-1.50	.13	.19	.41
Meaning recall – Time 1 fixation duration	0.23	[-0.03, 0.49]	0.13	1.26	[0.97, 1.63]	1.71	.09	.01	.22
Meaning recall – Time 2 fixation duration	-0.11	[-0.25, 0.03]	0.07	0.90	[0.78, 1.03]	-1.49	.14	<.01	.18
Meaning recall – Time 3 fixation duration	-0.09	[-0.25, 0.08]	0.08	0.92	[0.78, 1.08]	-1.03	.30	<.01	.16
Meaning recall – Cumulative fixation duration	0.10	[-0.17, 0.44]	0.15	1.11	[0.83, 1.48]	0.72	.47	.00	.22

4.9 Summary of the Results

This chapter has explored how task repetition influenced learners' processing of multimodal input and the acquisition of technical vocabulary. In addition, it has examined the association of multimodal processing with comprehension and vocabulary gains. Results of the eye-movement analyses showed that, in general, task repetition resulted in increased visual attention to the instructor but decreased attention to the diagram and target words. The number of integrative saccades between the instructor and the diagram declined during the second viewing and then increased in the third repetition. The stimulated recalls revealed that participants who performed the task repeatedly displayed notably fewer higher-level listening processes and used a greater proportion of strategies than those who did the task only once. Besides, the participants who had repeated the task twice reported engaging slightly more in reading the diagram labels but less in integrating information from different modalities than the other groups of participants. In terms of awareness of the target words, the participants in repeated viewing conditions made more comments about noticing specific form-related features and integrating target word information from different sources. Finally, participants' lecture comprehension, as measured by the number of correct idea units recalled, was found to be negatively related to both their visual attention to the instructor and the diagram. There was no significant association, however, between the participants' visual attention to the target words and their vocabulary gains.

This chapter has presented some additional findings in relation to participant-level and item-level differences that deserve attention. In terms of visual attention allocation during lecture viewing, a positive correlation was found between the size of the diagram AOI and fixation measures, whereas a negative correlation was observed with the AOI instance skip rate. In contrast, the instructor AOI showed a reversed pattern. What is more, participants' visual attention to the target words was found to be positively associated with larger AOI size, and those with higher vocabulary size tended to skip more target words. Also, nonverbal signals appeared to play a facilitating role in directing the participants' attention to the target words. Regarding vocabulary acquisition, the participants' L2 listening proficiency level, as measured by the CAE test, was a significant positive predictor of their immediate and delayed vocabulary gains. Interestingly, adjectives were linked to greater vocabulary gains than nouns.

Chapter 5 Discussion and Conclusion

In this final chapter, the results of the study are discussed in relation to the research questions and hypotheses proposed in Chapter 2, followed by a discussion of the theoretical, methodological, and pedagogical implications of the results obtained. The chapter concludes with a description of the limitations of the study and possible directions for future research.

5.1 Task Repetition and Attention to Lecture Content

The first research question explored: *To what extent does repeating a multimodal lecture-viewing task affect learners' visual attention to the lecture instructor and labeled diagrams, as reflected in their eye movements?* For this research question, I hypothesized that task repetition would lead to increased visual attention to the instructor and decreased visual attention to the diagram. The results showed that with repeated viewing of the lecture, the repetition group spent significantly more time looking at the instructor but less time reading the diagram. These results strongly confirm the first hypothesis. Drawing on Mayer's (2014c) cognitive theory of multimedia learning, learners probably needed to divide their visual attention between the instructor's presence and the labeled diagram, since all visual stimuli, including nonverbal, written verbal, and pictorial information, were processed in the visual channel with limited capacity. Therefore, during the initial viewing, participants might have paid more attention to the diagram, as written words could facilitate speech segmentation, and illustrations allowed them to quickly understand key knowledge elements (Schnotz & Wagner, 2018). During subsequent viewings, however, the role of the diagram might have been less important, considering that their semantic content was less clearly defined than that of the aural commentaries, and thus might not have supported learners in systematically constructing knowledge elements (Schnotz & Wagner, 2018).

Another possible reason for the participants' reduced visual attention to the diagram may be that longer fixations, as suggested in previous research (Jacob & Karn, 2003), indicated that viewers had difficulty extracting information from a display. Given that the diagram contained verbal and pictorial information representing unfamiliar knowledge elements, it is likely that the participants experienced comprehension difficulty and thus spent a longer time and fixed more frequently upon the elements during their first viewing. The decreased total fixation durations and counts in the subsequent viewings might reflect participants' increased familiarity with the diagram, regardless of whether or not they successfully understood the information presented. The stimulated recall comments additionally suggested that note-viewing might have contributed to the declined visual attention to the diagram. Specifically, participants who had engaged in the task more than once reported a higher frequency of reading or reorganizing their notes, leading to a reduced allocation of visual attention to the diagram.

On the other hand, the increase in learners' visual attention to the instructor during task repetition might have been due to the fact that the instructor's presence could not provide rich semantic information, causing the learners to shift their attention elsewhere during the initial viewing. When they became more familiar with the lecture content in the subsequent viewings, they might have been automatically attracted to the instructor's facial expressions and gestures (Beattie et al., 2010; Gullberg & Holmqvist, 2006). This interpretation is consistent with Suvorov's (2018) finding that visuals related to the lecturer were less frequently perceived by test-takers as meaningful and helpful for lecture comprehension than those related to the lecture content. This interpretation is further supported by the stimulated recall comments, which revealed that participants who had performed the task three times tended to pay more conscious attention to the instructor's nonverbal cues. For example, one participant described their viewing pattern, stating “I had already listened to this part twice, so I knew what she was talking about. I was naturally looking at her face while listening.”

Furthermore, the results uncovered a decline in the number of integrative saccades between the instructor and diagram from the first to the second viewing, followed by an increase from the second to the third viewing. In previous research (e.g., Pellicer-Sánchez et al., 2020; Pellicer-Sánchez et al., 2021), frequent transitions between L2 texts and pictures have been interpreted as indicating learners’ integration processes. In line with this reasoning, the high occurrence of switching between the instructor and the diagram during the first viewing might have been a sign of learners establishing connections between auditory (i.e., aural commentaries) and visual information (i.e., labeled diagram), as the instructor was the carrier of auditory information. This step, namely, coordinating multimedia elements, is key in multimodal comprehension (Mayer, 2014c), but may place high cognitive demands on learners, resulting in reduced attentional resources available to process specific linguistic and pictorial information. The opportunity for a second viewing might have allowed learners to focus on more specific bits of information after having familiarized themselves with the content during the initial performance. The stimulated recall comments provided additional support for this interpretation, revealing that the participants who had performed the task twice made notably fewer comments about integrating information from different input sources and more about using the strategies of focusing attention (e.g., focusing on listening to what they did not understand on the first viewing) and monitoring (e.g., checking and verifying information for consistency with what had been processed previously), as compared to those who performed the task once or three times.

As for the increased integrative saccades between the second and third viewings, one possible interpretation is that learners may have achieved at least a moderate level of familiarity with the lecture content after watching the lecture twice, enabling them to pay more attention to the instructor's social cues on the third viewing. Indeed, according to the stimulated recall data, participants who had performed the task three times made more reference to attending to the instructor's pointing gestures and eye gaze toward the diagram, which in turn directed their attention to the on-screen diagram. A quote from one of the participants illustrates this pattern: "I was looking at where she was pointing at. She pointed at this 'cell body', so I looked at the 'cell body'."

In addition, it is noteworthy that the majority of eye-movement measures did not yield significant differences between the second and third viewings. This might have been due to individual differences in participants' viewing patterns, as suggested by the relatively large standard deviations of the eye-movement measures at time 3 (see Table 24). For example, Participants 13 and 14 in the repetition group spent less than 1% of the time (calculated by dividing the total fixation duration to the instructor or the diagram by the total presentation time) looking at the instructor or the diagram during the third viewing, while Participants 1 and 29 looked at these areas for more than 25% of the time. These individual differences were also reflected in the stimulated recall comments, with some participants reporting reading and reorganizing their notes or even looking at irrelevant information on the screen (e.g., lecture background objects) during the third viewing, whereas others mentioned naturally looking at the instructor's face and the diagram as they had already taken notes. This finding is consistent with previous research (e.g., Ockey, 2007; Suvorov, 2015) that has found notable differences in how test-takers interacted with lecture videos.

Finally, an intriguing finding emerging from the data concerns the relationship between AOI size and learners' visual attention to the instructor and the diagram. Specifically, the size of the diagram AOI was positively correlated with fixation measures (i.e., duration and count), while a negative association was observed with the skip rate. However, an opposite pattern was found for the instructor AOI, with larger size being linked to decreased fixation durations and counts but increased skip rates. These findings may be attributed to the legibility of the diagram. As the camera zoomed in, both the instructor and diagram AOIs increased in size, but participants' attention might have been primarily drawn to the diagram because a larger size could facilitate easier reading of the hand-drawn illustrations and handwritten words, especially when the information was presented in a small area. This interpretation finds support in the stimulated recall data, where some participants reported having difficulty recognizing certain on-screen words and finding it easier to copy notes from the whiteboard when they were enlarged. As a consequence of the increased allocation of visual attention to the diagram AOI, the instructor AOI received less attention despite its increased size.

5.2 Task Repetition and Viewing Processes

The second research question asked: *To what extent does repeating a multimodal lecture-viewing task affect learners' multimodal listening processes, as reflected in their stimulated recall comments?* I posited that task repetition would have a considerable impact on the various processes underlying multimodal lecture comprehension and the way learners use listening/viewing strategies. The stimulated recall comments revealed that the majority of cognitive processes reported were related to L2 listening, with participants who had performed the task repeatedly mentioning a lower proportion of higher-level listening processes (i.e., meaning construction and discourse construction). These findings contradict those of previous studies (e.g., Field, 2015; Holzknecht, 2019), which found that double play elicited more reference to higher-level processes than single play. This inconsistency may be due to the differences in the nature of the listening tasks used across studies. In Field's (2015) and Holzknecht's (2019) research, participants were asked to complete multiple-choice and gap-filling questions that required listening for specific details, whereas the participants in the current study were instructed to watch and retell the lecture, so they might have engaged in more meaning-building processes during the initial viewing and then shifted their attention to lower-level processes (e.g., listening for precise sounds, words, and literal meaning of utterances) during the subsequent viewings. This inconsistent finding could also be attributed to participants' L2 proficiency levels. The participants in Field's (2015) study were pre-sessional students, presumably at a low intermediate proficiency level, but the participants in the current study were at an upper-intermediate to advanced level. As a result, the participants in Field's (2015) experiment might have encountered more challenges decoding L2 auditory input and matching what they heard to words, and subsequently reported relying more on lower-level processes in the first listening.

Despite this discrepancy, the results of stimulated recall comments also revealed that the participants used a wider range and a greater proportion of strategies during repeated viewing, which is in line with Holzknecht's (2019) findings. In terms of cognitive strategies, linguistic inferencing, prediction, and translating were adopted more frequently by learners in the initial task performance. A possible reason may be that these strategies were important for the participants to grasp the gist of the lecture when they were not familiar with the content. They had to infer meaning of technical words from definitions, anticipate key information, and use their knowledge of the L1 to make sense of what is heard. A stimulated recall participant who performed the task once, for example, mentioned that "I was listening to her explaining about its function and how it (*dendrite*) is distributed. I had to listen to her explanation because I did not know it in Mandarin." In addition, the participants who had performed repeatedly made more comments about using other cognitive strategies, namely, kinesic inferencing, elaboration, and organizing notes. This may be because the participants had sufficient time during repeated viewing to infer meaning from nonverbal social cues, elaborate on the missing information from outside the text, and organize information in ways that enhance comprehension. The strategy compensation (avoidance) was used across all three conditions, which supports the claim that all L2 listeners need to resort to compensatory strategies from time to time in order to construct meaning (Rost, 2011).

For metacognitive strategies, participants who had performed the task repeatedly reported more frequent use of the strategies of planning, focusing attention, and monitoring. A plausible explanation is that these are resource-intense strategies, and learners might not have been able to use them during the first viewing due to limited processing capacity. When participants were familiar with the lecture content in subsequent viewings, their attentional resources might have been sufficient to develop plans to overcome comprehension difficulty or to check and correct comprehension. Besides, only one participant reported using a socio-affective strategy (managing emotions) during the third viewing. It might be that the participants in the current study were familiar with the task type (i.e., listening to a lecture while taking notes) and completed it in a low-anxiety classroom situation, so they did not rely on this strategy as much as test-takers in a listening testing context. Finally, task-specific strategies emerged as an additional category of listening/viewing strategies during the coding. The participants who had performed the task twice mentioned using task-specific strategies notably more often, such as selecting important information and remembering pronunciation and the instructor's words, compared to the participants who had performed the task once or three times. This result suggests that once the participants had understood the lecture content after the initial viewing, they were more likely to focus on preparing for the free recall test that followed.

Turning now to reading comprehension, participants who had performed the task twice commented on graphemic input analyses (i.e., recognizing words in written form) slightly more than participants in the other two stimulated recall groups. When participants were exposed to verbal and pictorial input concurrently for the first time, they might not have been able to fully process the written words presented due to their limited attentional capacity. During the second viewing, however, more attentional resources might have been available to process the written words. When viewing for the third time, participants did not necessarily need to look at the on-screen words, as they might have become quite familiar with the words or have copied them down from the whiteboard into their notes. Along the same lines, the stimulated recall data showed that the participants tended to notice more target word features during the second viewing, which will be discussed in Section 5.6. In terms of visual comprehension, no notable differences in either the lower-level (i.e., identifying graphic displays in the diagram) or higher-level (i.e., mapping the graphic displays to form a mental model) processing were observed across the three groups, suggesting that pictorial information (i.e., diagram illustrations) might have served as an important source of input that participants could rely on to construct meaning throughout the repetition.

With regard to integrating multimodal L2 input, participants who had performed the task twice reported the lowest proportion of comments concerning perceptual (i.e., connecting words and graphical elements) and semantic processing (i.e., connecting propositional descriptions with structural characteristics of the mental model), compared with the other two groups of participants. This result is well aligned with the eye-movement measures that revealed a decrease in integrated saccades between the instructor and the diagram from the first to the second viewing, but an increase from the second to the third viewing. As discussed in Section 5.1, participants possibly needed to engage in integration processes to understand the lecture content during the initial viewing, thus might not have had sufficient attentional resources available to process specific written words and illustrations due to cognitive demands imposed by the task. Being able to view the lecture again might have diverted participants' attention to specific areas where they were having difficulty processing, such as a piece from the aural commentaries or spellings of certain words. When watching the lecture for the third time, they might have already been acquainted with the lecture content, so they could probably pay attention to the information conveyed in the diagram as well as the instructor's nonverbal social cues at the same time, resulting in more integrative transitions between these two areas.

Lastly, two meta-codes, note-viewing and processing of the instructor's social cues, emerged during the coding. Only participants who had performed the task repeatedly mentioned reading their notes, indicating that notes became an important source of input during repeated viewing. Furthermore, participants who had performed the task three times made more comments about the instructor's nonverbal communication cues. This trend was also captured in the eye-gaze recordings, showing that the amount of time participants spent looking at the instructor significantly increased during the third viewing period. This may indicate that the participants naturally tended to look at the instructor as the subject of the communication (Mayer, 2014c; Mayer et al., 2003) after they had gained sufficient familiarity with the lecture content from the previous viewings. Taken together, the second hypothesis received only partial confirmation, given that task repetition did not have a noticeable impact on learners' visual comprehension.

5.3 Relationship between Attention Allocation and Comprehension

The third research question examined: *To what extent does repeating a* *multimodal lecture-viewing task affect the relationship between learners' visual attention to the lecture instructor as well as the labeled diagrams and their lecture comprehension?* A nondirectional hypothesis was formed for this research question. The results demonstrated that the control group's visual attention to the instructor was negatively correlated with their lecture comprehension. Besides, the repetition group's visual attention to the instructor at times 2 and 3, as well as their attention to the diagram at time 3, was found to be negatively associated with lecture comprehension. In terms of the relationship between attention to L2 written texts and comprehension, Pellicer-Sánchez et al. (2021) found that the time spent on the text was negatively related to L1 readers' comprehension of text-related questions, but was positively related to L2 readers' comprehension. Pellicer-Sánchez and colleagues interpreted L1 readers' increased amount of attention to the text as an indicator of processing difficulties. Drawing on this interpretation, a larger amount of visual attention to the instructor might as well have reflected the control group's difficulties in processing L2 auditory input, as they were likely to rely more on nonverbal social cues to aid their lecture comprehension. Another possible interpretation of the negative association is that the instructor's presence might have led to poorer performance because it acted as a source of visual interference that drew learners' attention away from the content-related information (Colliot & Jamet, 2018). As for the repetition group, those who spent a longer time on the instructor during repeated viewing probably focused less on viewing and reorganizing their notes based on the auditory commentaries, which might have potentially facilitated understanding, storage, and retrieval of the lecture content.

Regarding the time allocated to the diagram, the findings of the current study are inconsistent with those of Pellicer-Sánchez et al. (2021), who found that longer time spent on pictures was positively linked to the comprehension of image-related questions for both L1 and L2 readers and was positively associated with L1 readers' comprehension of text-related questions. The authors argued that the positive relationships might have been due to L1 readers' well-developed reading skills, allowing them to spend less time on the texts and more time on deeper processing of pictures. On the contrary, in the current study, most of the semantic information was conveyed through the aural commentaries, and therefore, more time spent on the diagram might have indicated learners' difficulties in processing the auditory input. They might have relied on the written verbal and pictorial information to compensate for comprehension breakdowns. Alternatively, more time on the diagram might have also reflected difficulties in understanding the information being conveyed in the diagram, as discussed in Section 5.1. Thus, increased attention to the diagram during the third viewing might have taken up learners' cognitive resources used for processing auditory information, note-viewing, and note-reorganizing, leading to a detrimental effect on their performance in the free recall test.

5.4 Task Repetition and Vocabulary Acquisition

The fourth research question asked: *To what extent does repeating a multimodal lecture-viewing task affect learners' incidental acquisition of technical words, as measured by offline vocabulary tests?* The hypothesis formed for this research question assumed that participants who performed the task repeatedly would achieve greater vocabulary gains. As predicted, the results showed that participants benefited significantly from repeated viewing: the repetition group scored significantly higher than the control group on all tests (i.e., immediate and delayed form recognition, meaning recall, and meaning recognition tests). There are several possible explanations for the significantly greater development observed in the repetition group's lexical knowledge. First, task repetition probably helped learners add details to their developing representation of the target words. Specifically, participants might have understood the meaning of the target words during their initial viewing, and task repetition could have assisted them in noticing their form and linking the form to their meaning that they had previously extracted. This explanation is also supported by the stimulated recall data in that, participants in the repeated viewing conditions made more comments about paying conscious attention to form-related features of the target words than those who only performed the task once. For example, a participant who had performed the task twice reported that she understood the function of the cell body during the first viewing but did not notice the explanation that the cell body is also called soma. She then established the link between the word *soma* and its meaning in the second viewing.

Second, task repetition might have additionally facilitated the process of integrating lexical information from multiple input sources. It is possible that the participants experienced cognitive overload during the first viewing as they had to comprehend a lecture on an unfamiliar topic, especially when the lecture contained new technical words that were presented in real time using various types of L2 auditory and visual input, including spoken elaborations, written forms of the target words, diagram illustrations, and an instructor's nonverbal visual signals. Task repetition could potentially ease participants' cognitive demands posed by initial lecture viewing, and subsequently, promote the integration of the lexical information through the different modalities. This interpretation is also supported by the stimulated recall data: participants who had performed the task repeatedly reported with greater frequency that they relied on multiple input sources to understand the meaning of the target words.

Notably, the findings regarding vocabulary acquisition are consistent with those of Shintani (2012a), who reported positive effects of repeating a listening-and-do task on learners' gains in both receptive and productive knowledge on immediate and delayed post-tests. The findings, however, differ from those of Ellis and Chang (2016) and

Majuddin and Siyanova-Chanturia (2021), who observed only immediate gains in learners' lexical knowledge after repeated listening and viewing, respectively. One feature that distinguishes Shintani's (2012a) and the current study from the latter experiments is that the processing of target words was essential to task completion. In Shintani's study, participants had to point to pictures to show their understanding of the target words, and the participants in the current study were asked to take notes on a lecture in which the target words were technical terms representing key concepts and then leave a voice message. In contrast, Ellis and Chang's information-transfer tasks, as the researchers acknowledged, did not directly attract learners' attention to target words.

An alternative explanation of the repetition group's significant vocabulary gains draws on Robinson's Triadic Componential Framework (2001a, 2001b). According to the framework, note-taking created a dual-task situation that increased cognitive task complexity during the initial viewing, potentially diverting learners' attention away from comprehension and learning (Robinson, 2001a). A stimulated recall participant who had performed the task only once, for example, reported that she did not hear what the instructor said about a target word because of note-taking. The increased task complexity resulting from the secondary task (i.e., note-taking), however, might probably be reduced through task repetition. As learners familiarized themselves with the lecture content after the initial viewing, their attention could be redirected to linguistic aspects of the lecture that might have been overlooked due to note-taking. In addition, the need to take notes might have encouraged participants to engage in deeper processing of the target words in subsequent viewing (Di Vesta & Gray, 1973). Notes also provided a physical record of lecture content from which information could be recalled and rehearsed when learners read them during repeated viewing (Di Vesta & Gray, 1973). This might have also increased the chance of deeper learning of the target words. Note-taking and note-viewing, in turn, likely led to better integration of new and prior lexical knowledge, ultimately resulting in greater vocabulary retention (Craik & Lockhart, 1972; Gablasova, 2014; VanPatten, 2004; VanPatten et al., 2004).

In addition to different task demands, the inconsistent findings might be attributed to the presence of rich context. While Ellis and Chang (2016) and Majuddin and Siyanova-Chanturia (2021) used target words whose meaning was provided implicitly in the context, Shintani's (2012a) and the current study presented target words with explicit spoken elaborations and pictures demonstrating underlying concepts. Therefore, participants in the latter experiments were more likely to notice the form of target words and infer their meaning from the input. Task repetition could have further increased the probability that learners focus on the meaning of target words that were only partially understood during the first viewing, add details to already stored meaning, and construct form-meaning associations. Similarly, this initial deeper processing of target words might also explain the greater retention of vocabulary (Craik & Lockhart, 1972).

An additional finding worth noting is that the POS of target words appeared to be a significant predictor of participants' vocabulary gains, with adjectives being linked to greater immediate and delayed meaning recognition. Although the exact ranking of the POS in terms of acquisition difficulty is not clear, Laufer (1990, p. 298) summarized a general trend: "It is sometimes argued that certain grammatical categories are more difficult to learn than others. Nouns seem to be the easiest; adverbs – the most difficult; verbs and adjectives – somewhere in between." This discrepancy in findings might have been due to the abstractness of the target words. The adjectives used in this study (i.e., *mesopic*, *photopic*, *scotopic*) represent different light levels that are perceivable in the real world, whereas the nouns (i.e., *arachnoid*, *axon*, *dendrite*, *dura*, *meninges*, *pia*, *soma*, and *synapse*) are names for important parts of the nervous system which are more difficult to conceptualize. From this perspective, this result corroborates previous findings that abstractness affects vocabulary acquisition (Gablasova, 2014; Laufer, 1997).

Apart from an item-level property of the target words, a participant-level difference also emerged as a significant predictor of learners' vocabulary learning: Higher CAE scores were related to greater immediate and delayed meaning recognition, as well as delayed meaning recall. A plausible explanation for this might be that less competent listeners were likely to experience difficulties in extracting meanings of the target words and subsequently were unable to form associations between the target word and its elaboration. They probably also spent more attentional resources on lower-level processes (i.e., input decoding, lexical search, and parsing), thus having less mental capacity available for noticing form-meaning connections. In contrast, these lower-level processes might be highly automatic for competent learners (Field, 2013). Therefore, more of their attentional resources might have been freed for higher-level processes (i.e., meaning and discourse construction), as well as noticing and consolidating form-meaning associations.

5.5 Task Repetition and Attention to Target Vocabulary

RQ5 explored: *To what extent does repeating a multimodal lecture-viewing task affect learners' visual attention to technical words, as reflected in their eye movements?* The hypothesis for this question posited that participants' visual attention to the technical words would decline during repeated task performance. Within the repetition group, the participants' visual attention to the target items indeed significantly decreased with repeated task engagement, providing strong confirmation for the hypothesis. This is in line with findings of previous research where readers spent significantly less time processing the unknown words after repeated encounters with the words (e.g., Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016). Drawing on the claim that the development of mental representation of new words as well as fluent access to the words requires repeated exposure (Wesche et al., 2010), this downward pattern was likely to be related to learners' more fluent access to lexical knowledge after building form-meaning links in the initial viewing. Alternatively, the finding can probably be attributed to learners' increased familiarity with the visuographic features of target words, whether or not learners made form-meaning connections (Pellicer-Sánchez, 2016).

In the current study, the downward trend might have been further enhanced by the multimodal nature of the input. Following Mayer's (2014c) cognitive theory of multimedia learning, learners likely had to split their attention between different types of visual information, all being processed in the same limited capacity visual channel. As a result, participants might have prioritized paying attention to the written form of the target words during the first viewing, given that they labeled the important parts of the diagrams closely related to task completion. In line with this reasoning, the stimulated recall comments suggested that participants were likely to allocate more of their conscious attention to illustrations or nonverbal visual signals to aid the understanding of the target words during subsequent viewing. For example, a stimulated recall participant who had performed the task twice reported paying attention to the instructor's body language to understand the function of *dura* because she knew from the first viewing that some words in the instructor's oral explanation would hinder her understanding of the content. The stimulated recall comments additionally revealed that the notes participants had taken during the initial viewing served as a key source of input during task repetition, further explaining the declined visual attention to the on-screen presentation of the target words. Some stimulated recall participants recalled that they had read or reorganized their notes during repeated viewing in order to memorize the target words or to check if the lexical information extracted from repeated viewing was consistent with what they had obtained earlier.

Additionally, the results revealed that larger AOI size and longer item length were linked to greater visual attention to the target words. This is in line with previous findings that students without prior knowledge of a subject tended to pay more attention to visually salient features (e.g., Lowe, 1999, 2003). More non-verbal signals (i.e., the instructor's gestures of pointing to a target item) were related to increased visual attention as well, providing further support for the signaling principle, that is, attention-directing cues in a multimedia message enhance deep learning (van Gog, 2014). This also accords with the findings of Suvorov's (2018) study that the instructor's movements, including pointing at the visual aids, were most commonly reported by test-takers as useful to attract their attention and facilitate comprehension. Interestingly, participants' vocabulary size emerged as a significant predictor of their visual attention, with larger vocabulary size linked to an increased skip rate of the target items. One explanation for this may be that those with a larger vocabulary size probably relied less on the written form of the target words to aid their understanding of the lecture content.

5.6 Task Repetition and Awareness of Target Vocabulary

The sixth research question asked: *To what extent does repeating a multimodal lecture-viewing task affect learners' awareness of technical words, as reflected in their stimulated recall comments?* It was postulated that participants' level of awareness of the target words would increase during task repetition. As expected, stimulated recall data showed that the participants who had performed the task repeatedly reported more comments associated with noticing specific aspects of the target words. This pattern is in line with the results of the vocabulary post-tests, which demonstrated superior gains by the repetition group. Furthermore, it is consistent with the results of the eye-movement data that found learners' visual attention to the target words reduced with repeated exposure. It is probably, as discussed earlier, due to participants' increased familiarity with the target items.

The stimulated recall data also revealed how task repetition affected the source of participants' awareness of the target words. Diagram labels and aural commentaries constituted the main source of noticing and understanding of target items regardless of task repetition. Interestingly, aural commentaries seemed to be the primary source of input for learners to derive the meaning of the target words, rather than the diagram illustrations presented in the lecture. This might be due to the fact that the simple hand-drawn illustrations did not provide a sufficient amount of information for learners to infer meaning of the technical words. Nonetheless, as discussed in Section 5.5, participants who performed the task repeatedly mentioned more frequently paying conscious attention to the target words in their notes and to various input sources used to introduce the technical words, including written forms of the words, their spoken elaborations, and the instructor's nonverbal communication cues. These insights emerging from the stimulated protocols provide an alternative interpretation for why a decrease was observed in the visual attention to the on-screen presentation of the target words, which was captured in the eye-gaze recordings.

5.7 Relationship between Attention Allocation and Learning

The final research question asked: *To what extent does repeating a multimodal lecture-viewing task affect the relationship between learners' visual attention to technical words and their vocabulary acquisition?* A non-directional hypothesis was proposed for this research question. In contrast to previous reading studies (e.g., Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016) and viewing studies (e.g., Wang & Pellicer-Sánchez, 2021), the results showed that neither group's visual attention to the target words predicted their vocabulary gains in either receptive or productive knowledge of the target words. Drawing on the stimulated recall comments, a possible explanation may be that the aural commentaries, rather than the visual information, were the main source of input for learners to understand the meaning of the target words. In addition, learners might have also prioritized remembering the oral rather than the written form of the target words, given that their subsequent task involved oral production (i.e., leaving a voice message to a friend). What is more, learners had the opportunity to write down the target words and read their notes during repeated viewing, providing a further reason why they might have perceived remembering written forms as less important. Also, the amount of visual attention allocated to the target words might not fully reflect the cognitive effort learners spent on processing the target words since notes served as an external input source.

It should also be noted that the distractors used in the form recognition test were not taken into consideration when analyzing the relationship between the eye-movement indices and immediate and delayed form recognition gains because they were not presented in the lecture. After excluding the distractors, the variance between the control and the repetition groups' form recognition scores was relatively small (immediate form recognition: control group, M = 8.96, SD = 1.35, 95% CI [8.46, 9.47], repetition group, M = 9.5 SD = 1.22, 95% CI [9.04, 9.96]; delayed form recognition: control group, M = 10.07, SD = 1.57, 95% CI [9.48, 10.65], repetition group, M = 10.37, SD = 1.00, 95% CI [10.00, 10.74]), and relatively high mean scores across groups can be observed, suggesting a potential ceiling effect. This may partly explain the nonsignificant results between learners' fixation duration on the target words and form recognition scores.

5.8 Implications of the Study

Having discussed the findings of this research in detail, this section presents the broader implications of this study for theory, methodology, and practice in the field of

SLA.

5.8.1 Theoretical Implications

This study has revealed several novel theoretical insights. First, it contributes to our understanding of the theoretical construct underlying video-based academic listening. In this study, learners paid a fair amount of visual attention to the instructor's social cues as well as the labeled diagram and frequently reported utilizing these visual stimuli, suggesting that visual information plays an important role in multimodal lecture comprehension. This finding, therefore, supports the calls from L2 researchers to expand the academic listening construct by including the ability to understand visual information (e.g., Ockey, 2007; Suvorov, 2015), given that visual input is included in most real-world academic listening situations in conjunction with the spoken language (Wagner & Ockey, 2018).

Second, the results of the current research contribute to our understanding of the effect of task repetition on learners' processes involved in video-based lecture comprehension. Although existing work (e.g., Bygate, 1996) has theorized the effects of task repetition on speaking processes based on L2 speech production models (e.g., Levelt, 1989), no prediction has been made regarding how task repetition may affect learners' cognitive processes during L2 viewing. To fill this gap, this study has formulated an integrated model drawing on models of L2 listening processes (Field, 2013), listening strategies (Vandergrift & Goh, 2012), and multimedia learning (Schnotz, 2014). The integrated model conceptualized multimodal lecture comprehension as the cognitive and strategic processing behavior that learners engage in to understand multimodal input. Except for visual comprehension, task repetition appeared to affect various processes underlying multimodal lecture comprehension, that is, listening comprehension, lower-level reading comprehension, integration of information from different modalities, and strategy use. The stimulated recall comments lent additional support to Mayer's (2014c) cognitive theory of multimedia learning that successful multimodal comprehension requires effective selection and organization of auditory, written, and pictorial information, integration of information from different sources, and appropriate use of strategies. Therefore, these cognitive models appear as useful theoretical starting points for future work exploring processes underlying L2 multimodal comprehension.

Third, the results yielded in this study also shed light on how learners' viewing behavior affects their lecture comprehension. Although the findings may appear to differ from those of previous L2 reading research (e.g., Pellicer-Sánchez et al.,2020; Pellicer-Sánchez et al.,2021), which have found that visual attention to pictures was positively associated with reading comprehension, it is important to note that this difference may reflect the complexities of L2 viewing: learners need to process spoken language almost immediately, thus longer processing time on visual aids (nonverbal signals, written verbal and pictorial information) might indicate difficulties in understanding L2 auditory information regardless of whether or not learners repeated the task. The current study, therefore, has contributed to the growing body of knowledge about the complex relationship between attention to visual stimuli and L2 viewing comprehension.

Fourth, the present study has provided valuable insights into the impact of task repetition on learners' processes underlying vocabulary acquisition from multimodal L2 input. While previous studies have investigated learners' attentional processes involved in vocabulary acquisition from repeated exposures (e.g., Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016), no investigation has explored how repeated task engagement affects learners' awareness of target lexical forms. This study contributed to this line of research, drawing upon Schmidt's (2001) Noticing Hypothesis (Schmidt, 2001) and Hegarty's (2014) information processing model. The findings of the current study echo the prevailing view in SLA that researchers should untangle the constructs of attention and awareness both conceptually and empirically (e.g., Godfroid et al., 2010; Godfroid et al, 2013), as the findings showed that learners' visual attention to the target words declined, but their awareness of the target words increased with repeated viewing. It was also found that the diagram labels and aural commentaries constituted the main source of noticing and understanding of the target items, respectively; this enriches our understanding of the source of learners' awareness of target words. As such, Schmidt's (2001) Noticing Hypothesis and Hegarty's model (2014) can provide a solid theoretical foundation for further research investigating vocabulary acquisition from repeated exposure to multimodal L2 input.

In addition, the findings of this study support Skehan's (1998) limited capacity model which assumes that task repetition can ease learners' cognitive demands as regards content and subsequently direct their attention to linguistic forms. The stimulated recall comments confirmed that the learners experienced cognitive overload when constructing meaning from information presented in multiple modalities during the first viewing, and they were able to notice more specific aspects of the target words (e.g., orthographic form, phonological form, GPC, and POS) during subsequent viewing when they were more familiar with the lecture content.

5.8.2 Methodological Implications

The findings of this study also have a number of methodological implications. First, as compared to static AOIs used in previous studies (e.g., Suvorov, 2015), dynamic AOIs allow researchers to obtain richer information about how L2 learners interact with lecture videos, as dynamic AOIs can be set to concurrently change their shape and position to follow an object on the screen, thus enabling researchers to track learners' attentional allocation to a moving image (Batty, 2020). Besides, dynamic AOIs can be created as irregularly shaped areas, providing more precise measurements of where and when learners' visual attention is directed than rectangular or circular AOIs. Thus, dynamic AOIs can be used in future viewing studies to provide better insights into learners' attentional processes during L2 viewing, as well as how visual processes can be influenced by other factors, such as the size and presentation duration of visual stimuli.

Another methodological contribution of this study is the triangulation of multiple data sources, including eye movements, stimulated recalls, offline vocabulary post-tests, and free recall protocols, to analyze learners' processes underlying multimodal lecture comprehension and incidental vocabulary learning. Specifically, while the vocabulary post-test and free recall protocols examined the outcomes of comprehension and vocabulary acquisition, eye-movement data provided valuable information about learners' real-time attentional processing of multimodal L2 input, and the stimulated recalls additionally revealed learners' conscious cognitive processes that could not be reflected in eye movements. The inherent weakness of each method, therefore, was reduced, thus maximizing the internal and external validity of research (Dörnyei, 2007). The results obtained using this triangulation approach also supports the calls for disentangling the constructs of attention and awareness both at conceptual and empirical levels, as discussed previously. In future research, researchers should consider using the triangulation of quantitative and qualitative data sources to paint a fuller picture of learners' viewing behavior (eye-tracking) and thought processes (stimulated recall) during viewing, as well as the way attentional and awareness relate to learning outcomes.

5.8.3 Pedagogical Implications

Moving on to pedagogy, an important implication of this research is the advocacy of using meaningful tasks and task repetition to facilitate L2 processing and learning. Along with previous research (e.g., Joe, 1995, 1998), the current study has provided empirical support that task demands govern the quality and amount of attention to L2 features (Huckin & Coady, 1999). In particular, the study proves that the lecture-viewing task and the announced free recall test encouraged learners to process the lecture input with a clear purpose, prompting them to notice the target words and use various task-specific strategies (e.g., listening for gist, focusing on specific form-related features, and synthesizing information from their notes), and subsequently contributed to the learning and retention of vocabulary knowledge. Thus, teachers may consider incorporating pedagogical tasks in language learning classrooms to guide learners' attention to L2 features and help them develop cognitive and linguistic skills that can be applied to real-world communication.

In addition, the results revealed that repeated viewing led to lower reliance on higher-level listening processes. This insight may help teachers design appropriate classroom instructions that align with natural cognitive processes involved in repeated viewing. For instance, teachers can design listening activities that start with listening for the gist, allowing learners to grasp the overall meaning of the content. This can be followed by listening for specific information which aims to enhance learners' skills in lower-level processing. By sequencing activities in this manner, teachers may optimize learners' engagement and comprehension. Moreover, the results showed that repeating a lecture-viewing task can help learners cope with cognitive demands imposed by having to process multimodal L2 input while taking notes, thereby enabling learners to better comprehend the lecture content and allocate more attention to L2 forms. Repeated engagement also significantly enhanced the acquisition and retention of technical vocabulary. These findings suggest that task repetition is a useful task implementation factor for language learners, thus teachers can encourage students to view short audio-visual materials repeatedly in order to optimize L2 learning. The practical implications of the results may as well extend to university-level policy makers, who can consider implementing measures that enable students to easily access lecture recordings for repeated viewing. By adopting such policies, there is a promising opportunity to facilitate the academic achievements and linguistic development of students in L2-medium education.

Finally, the findings also uncovered that the instructor's pointing gestures play a facilitating role in guiding learners' attention to word forms that were referring to. It may also be useful for teachers to consciously use nonverbal signals (e.g., directional gaze or pointing gestures) to highlight the target linguistic forms that they want their students to master.

5.9 Limitations and Directions for Future Research

In interpreting the potential contributions of this research, it is also important to consider the limitations of the study. These relate to the (a) generalizability of the findings, (b) selection of materials and instruments, and (c) data collection and analyses. Regarding generalizability, the participants were made aware of the free recall test when given the instruction for the lecture-viewing task, which might have affected their cognitive processes and strategy use during viewing (Nguyen & Boers, 2018). Drawing on Swain's (2005) output hypothesis, which claims that "the act of producing language (speaking or writing) constitutes, under certain circumstances, part of the process of second language learning" (p. 471), the announcement of the free recall test potentially encouraged learners to attend to keywords related to the main ideas presented and become prepared to use the newly encountered words during the performance of the lecture-viewing task. As the stimulated recall comments demonstrated, participants in all conditions reported using task-specific strategies, such as selecting important information and remembering phrases and pronunciation of keywords in order to retell the lecture.

Accordingly, the findings may not be generalizable to tasks focusing solely on lecture viewing. In addition, this study only employed short lecture videos on neurobiology as the viewing materials. Future research using audio-visual academic materials with different length or topics are therefore warranted.

In terms of the selection of materials and target words, a limitation originated from the number of target items and their legibility. Considering that the participants had to watch the lecture videos repeatedly, it was not feasible to use extensive videos which contained potentially more target lexical items (Majuddin & Siyanova-Chanturia, 2020). The handwriting quality of target words could not be controlled either due to the authentic nature of the video prompts, which might have led to relatively longer fixations on the target words. To provide a fuller picture of vocabulary learning processes in a multimedia learning context, it would appear worthwhile for future eye-tracking research to investigate how word characteristics (e.g., frequency of occurrence) influence the effectiveness of incidental vocabulary acquisition from viewing. A further weakness relates to the effect of the free recall test on learners' retention of the target words, as note-viewing and retelling provided extra learning opportunities that could strengthen memory of the words (Nguyen & Boers, 2018). It is difficult, however, to assess learners' comprehension without using any comprehension tests (Serrano & Huang, 2018).

As for data collection and analyses, although this study provided rich insights into L2 learners' processing of multimodal L2 input and acquisition of technical vocabulary through triangulating different types of data, this analysis is not without its limitations. First, the findings would have been more fine-tuned if the analyses of participants' notes were included, which might have provided further insights into the lower-level processes in which participants engaged but might not have become aware of or were unable to report (Rukthong & Brunfaut, 2020). In future research, the incorporation of participants' notes (if there is any) would probably help investigate the L2 listening/viewing process in an academic setting. Second, despite the use of a 25 mm remote lens had improved the efficiency of the eye-tracking system in detecting pupils (it is by default configured to use a 16 mm remote lens), the system had to re-capture participants' pupils each time they shifted their gaze from their notes to the computer screen. This might have resulted in some measurement errors or delays in capturing participants' eye movements accurately. Given that the effect of note-taking was not taken into consideration during the analysis of the eye-movement data (i.e., calculating the percentage of track loss), the results for participants' viewing behavior in this study should be interpreted with caution.

Another limitation concerns the complexities associated with dynamic AOI. First, even though the dynamic AOIs were created to move simultaneously with the targets (i.e., the instructor and the diagram), the location and duration of fixations are completely independent of the location and duration of individual instances of an AOI. For example, the instructor appeared on the screen at the beginning of the video for 100 milliseconds and then started to move to the left of the screen at time 101 milliseconds. Accordingly, two instances of the instructor AOI were created: one that lasted for 100 milliseconds, and the other that aligned over the instructor's moving position. A participant might have looked at the instructor's face from the beginning and continued fixing upon the same spot until time 130 milliseconds, after the instructor AOI had moved away. Therefore, the total duration of this fixation on the instructor was 130 milliseconds, and only a proportion of the fixation duration (100/130 milliseconds) "belongs" to the first instance of the instructor AOI. However, the current study had to adopt a traditional fixation-based approach, that, is allocating an entire fixation duration to an AOI (in this example, fixation duration was calculated as 130 milliseconds instead of 100 milliseconds), given that it is not practical to segment each fixation for every participant due to the video length. It may be more appropriate for future studies to use a Perform Sample Based Calculation provided by the EyeLink Data Viewer (2019) software, which allows fractions of fixation duration to be added to dynamic AOIs.

Furthermore, following previous research (e.g., Bisson et al., 2014), the dynamic AOIs created for the areas around the instructor, diagrams, and target words were slightly larger than the exact borders to absorb a small amount of drift. This, however, might have undermined the precision of the eye-tracking data obtained in the current study, given that the targets occasionally overlapped with each other (e.g., when the instructor put her hands on the diagram on the whiteboard). Future studies could consider using software tools that support the automatic generation of dynamic AOIs to reduce the time and effort required to analyze dynamic stimuli while maintaining a high degree of accuracy.

Another limitation that deserves attention is that the current study did not directly compare the allocation of attention to the instructor and the diagram AOIs. Consequently, decreased attention to the diagram AOI during one exposure may not necessarily mean that it attracted less learners' attention than the instructor AOI. Percentage eye-movement measures (e.g., dwell time percentage and fixation percentage) could be used in future studies to address this limitation. By comparing learners' visual attention to different parts of the lecture, researchers could identify which elements are more salient and engaging for learners. Another shortcoming of this research lies in the fact that diagram illustrations and written labels were included in one dynamic AOI, although they are different types of visual information. Nonetheless, it was impractical to create separate dynamic AOIs for these two types of visual stimuli as they were presented very close together (see Figures 11, 12, and 13). Thus, the interpretation of how learners integrated auditory information with pictorial and written verbal information was more tentative. The absence of an AOI for other context-related information (e.g., background objects) is an additional weakness of this study. Including such an AOI would have provided a fuller picture of learners' attention allocation during lecture viewing, as stimulated recall participants mentioned looking at room decorations when they were quite familiar with the lecture content during the third viewing.

The use of stimulated recall interviews also has its inherent limitations. The participants could not report all of their conscious processes during task performance, and the recall protocols might have become less accurate as the number of times they performed the task increased (Gass & Mackey, 2016). For those who did the same task repeatedly, they might have reported something that did not happen during their last performance, despite the fact that the participants were explicitly instructed to recall their last performance. In addition, the participants could only report their conscious awareness of the target words and the source of awareness. When a participant reported hearing a target word, this does not necessarily mean that they did not simultaneously look at the word presented on the screen or in their notes. Thus, the stimulated recall results should also be treated with caution.

Besides addressing the limitations of the current study, it would be interesting for future research to compare the effects of distributed and massed repetitions on learners' processing and acquisition of technical words. The spacing effects have been widely investigated in L2 reading studies (e.g., Koval, 2019), and the insights gained from the investigation could help inform the development of more efficient and effective L2 learning methods. Future research is also needed to look into the role of multimodal academic materials that provide navigational control for learners to replay, pause, and move forward, given that language learners today are in full control of various academic materials (e.g., lecture recordings and online courses) in everyday listening/viewing situations. Another relevant area for further research is the use of neurophysiological data to explore the cognitive mechanisms underlying L2 viewing and learning from multimodal input (Révész & Shi, 2023; Suvorov, 2018). For example, electroencephalography (EEG) measures can be employed to assess learners' attentional processes during L2 viewing and learning. Last but not least, future research would also benefit from the investigation of the relationship between individual differences (e.g., language learning aptitude and working memory capacity) and L2 development, which allows researchers to infer learners' cognitive processes during L2 viewing.

5.10 Concluding Remarks

To conclude, the current study has utilized a data triangulation approach, combining eye-tracking, stimulated recall, free recall test, and vocabulary post-tests, to provide a comprehensive examination of the effects of task repetition on learners' processing of multimodal L2 input and acquisition of technical vocabulary. The study found that repeating a video-based lecture-viewing task led to increased visual attention to the instructor but decreased visual attention to the diagram. Interestingly, the number of integrative transitions between the instructor and the diagram declined from the first to the second viewing, but increased from the second to the third viewing. These trends were also reported by the stimulated recall participants, who additionally described engaging in fewer higher-level listening processes and using a greater proportion of strategies during repeated viewing. The amount of attention paid to either the instructor or the diagram was found to be negatively associated with learners' lecture comprehension. The study has also revealed a facilitating effect of task repetition on learners' acquisition of technical words through lecture viewing: task repetition led to greater acquisition of technical words embedded in the lecture. The amount of visual attention to the technical words reduced with repeated viewing, suggesting increasingly fluent access to the words. However, it did not predict learners' vocabulary gains regardless of whether or not they completed the task repeatedly.

Overall, the findings suggested that task repetition helped learners deal with cognitive demands imposed by having to process L2 multimodal input in real time while taking notes. Due to repeated performance, learners could direct more conscious attention to L2 form-related features and integrate multiple sources of information, and this in turn facilitated form-meaning associations and subsequent learning. Having discussed the findings of the study, it is clear that this research has important theoretical and methodological, as well as pedagogical, implications. First, the results of the investigation provide novel insights into the processes underlying video-based lecture comprehension and vocabulary acquisition from repeated exposure to multimodal L2 input. The methodological novelty of the current research lies in triangulating eye-tracking with stimulated recall data, allowing for obtaining a richer and fuller picture of L2 learners' cognitive processes while processing multimodal L2 input. This combination is just beginning to be employed in the wider field of L2 research, and it has also been found helpful to gain information about both attention and awareness (Godforid et al., 2013). Finally, the findings of the study are of pedagogical value. Considering that language learners today, due to the surge of online learning during the COVID-19 pandemic, are regularly exposed to video-based lectures which they can watch repeatedly, the insights gained from this study may help teachers create and implement teaching materials in ways that facilitate the processing of multimodal L2 input and development in linguistic knowledge.

References

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research*, *7*, 347–376. https://doi.org/10.1191/1362168803lr127oa

Altarriba J., Kroll J., Sholl A., & Rayner K. (1996) The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, *24*, 477–92. https://doi.org/10.3758/BF03200936

Anderson, J. R. (1985). *Cognitive psychology and its implications*. Worth.

Arndt, J., Schüler, A., & Scheiter, K. (2015). Text–picture integration: How delayed testing moderates recognition of pictorial information in multimedia learning. *Applied Cognitive Psychology*, *29*, 702–712. https://doi.org/10.1002/acp.3154

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

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 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 Science*, *4*, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Badger, R., & Yan, X. (2012). The use of tactics and strategies by Chinese students in the listening component of IELTS. In P. Thompson (Ed.), *IELTS Research Reports Volume 9* (pp. 67–96), British Council/IELTS Australia. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_rr_volume09_report2.ashx

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2022). MuMIn: Multi-model inference. R package (Version 1.47.1). Retrieved from https://cran.r-project.org/web/packages/MuMIn/index.html

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Batty, A. O. (2020). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, *38*, 511–535. https://doi.org/10.1177/0265532220951504

Beattie, G., MacLeod, T., & Walker, P. (2010). Allocating visual attention to dynamic facial expressions and gestures. *Cognition*, *115*, 357–364. https://doi.org/10.1016/j.cognition.2010.02.008.

Bisson, M. J., Van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*, 399–418. https://doi.org/10.1017/S0142716412000434

Blom, E., Paradis, J., & Duncan, T. S. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular–s in child L2 English. *Language Learning*, *62*, 965–994. https://doi.org/10.1111/j.1467-9922.2012.00715.x

Bowles, M. A. (2008). Task type and reactivity of verbal reports in SLA: A first look at a L2 task other than reading. *Studies in Second Language Acquisition*, *30*, 359–387. https://doi.org/10.1017/S0272263108080492

Bowles, M. A. (2010). Concurrent verbal reports in second language acquisition research. *Annual Review of Applied Linguistics*, *30*, 111–127.

https://doi.org/10.1017/S0267190510000036

Bowles, M. A., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition*, *27*, 415–440. https://doi.org/10.1017/S0272263105050187

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Palgrave Macmillan.

Bygate, M. (2018). Introduction. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 1–25). John Benjamins.

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks in second language learning, teaching, and testing.* Pearson Education.

Bygate, M. & Samuda, V. (2005) Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). John Benjamins.

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

Chaudron, C. (1982). Vocabulary elaboration in teachers' speech to L2 learners. *Studies in Second Language Acquisition*, *4*, 170–180. https://doi.org/10.1017/S027226310000440X

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.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

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

Colliot, T., & Jamet, E. (2018). Understanding the effects of a teacher video on learning from a multimedia document: an eye-tracking study. *Educational Technology Research and Development*, *66*, 1415–1433. http://doi.org/10.1007/s11423-018-9594-x

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, *32*, 453–467. http://doi.org/10.1177/0267658316637401

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

Coxhead, A. (2018). *Vocabulary and English for specific purposes research: Quantitative and qualitative perspectives*. Routledge.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. 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*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Culter, A. & Clifton, C. (1999). Comprehending spoken language: a blue print of the listener. In C. M. Brown, & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 123–166), Oxford University Press.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*, 369–382. https://doi.org/10.1177/0267658312443651

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*, 436–459. https://doi.org/10.1002/tesq.552

Dang, T. N. Y., Lu, C., & Webb, S. (2021). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*, *44*, 708–736. https://doi.org/10.1017/S0272263121000474

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, *21*, 113–140. https://doi.org/10.1007/s10648-009-9098-7

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*, 278–287. https://doi.org/10.1037/h0034589

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). Cambridge University Press.

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, 88, 102166. https://doi.org/10.1016/j.system.2019.102166

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness*, *13*, 243–264. https://doi.org/10.1080/09658410408668810

Egi, T. (2008). Investigating stimulated recall as a cognitive measure: Reactivity and verbal reports in SLA research methodology. *Language Awareness*, *17*, 212–228. https://doi.org/10.1080/09658410802146859

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*, 341–366. https://doi.org/10.1017/S0272263117000109

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

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

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

Ellis, R. (2009). Task-based language teaching: sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*, 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Ellis, R., & Chang, L. (2016). The effects of inference-training and text repetition on Chinese learners' incidental Vocabulary acquisition while listening. *Chinese Journal of Applied Linguistics*, *39*, 137–152. https://doi.org/10.1515/cjal-2016-0009

Ericsson, K. A., & Simon, H. A. (1987). Verbal reports on thinking. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 24–53). Multilingual Matters.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. MIT Press.

Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, *30*, 198–208. https://doi.org/10.1016/j.esp.2011.01.001

Eyelink 1000 plus [Apparatus and software]. (2016). Mississauga, Ontario, Canada: SR

Research. Retrieved from https://www.sr-research.com/eyelink1000plus.html

EyeLink Data Viewer (Version 4.1.1) [Apparatus and software]. (2019). Mississauga, Ontario, Canada: SR Research. Retrieved from https://www.sr-research.com/data-viewer/

EyeLink Experiment Builder (Version 1.10.165) [Apparatus and software]. (2011). Mississauga, Ontario, Canada: SR Research. Retrieved from https://www.sr-research.com/experiment-builder/

Field, J. (2008). *Listening in the language classroom*. Cambridge 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

Field, J. (2015). *The effects of single and double play upon test outcomes and cognitive processing*. ARAGs Research Reports Online. The British Council. Retrieved from https://www.britishcouncil.org/sites/default/files/field_layout.pdf

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Fitzpatrick, T. (2013). Word associations. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 6193–6199). Blackwell.

Flowerdew, J. (1992). Definitions in science lectures. *Applied Linguistics*, *13*, 202–221. https://doi.org/10.1093/applin/13.2.202

Fox J. & Weisberg S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications.

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

Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures?. *Journal of Memory and Language*, *60*, 20–35. https://doi.org/10.1016/j.jml.2008.07.003

Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research* (2nd ed.). Routledge.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, *49*, 549–581. https://doi.org/10.1111/0023-8333.00102

Georgiou, S. (2012). Reviewing the puzzle of CLIL. *ELT journal*, *66*, 495-504. https://doi.org/10.1093/elt/ccs047

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, *19*, 133–167. https://doi.org/10.1191/0265532202lt225oa

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*, 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483-517. https://doi.org/10.1017/S0272263113000119

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing

Hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition* (pp. 169–197). John Benjamins.

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movementdata, verbal reports, and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). University of Hawai‘i.

Gruba, P. (1997). The role of video media in listening assessment. *System*, *25*, 335–345. https://doi.org/10.1016/S0346-251X(97)00026-2

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, *14*, 53–82. https://doi.org/10.1075/pc.14.1.05gul

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*, 339–353. https://doi.org/10.1037/1082-989X.3.3.339

Harding, L. (2011). *Accent and listening assessment*. Peter Lang.

Harrison X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, *2*, e616. https://doi.org/10.7717/peerj.616

Hatami, S. (2017). The differential impact of reading and listening on L2 incidental acquisition of different dimensions of word knowledge. *Reading in a Foreign Language*, *29*, 61–85. Retrieved from http://hdl.handle.net/10125/66728

Hegarty, M. (2014) Multimedia learning and the development of mental models. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 673–

701). Cambridge University Press.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717–742. https://doi.org/10.1006/jmla.1993.1036

Holzknecht. F. (2019). *Double play in listening assessment* [Unpublished doctoral dissertation]. Lancaster University. Retrieved from https://www.proquest.com/dissertations-theses/double-play-listening-assessment/docview/2472069878/se-2

Holzknecht, F., Eberharter, K., Kremmel, B., Zehentner, M., McCray, G., Konrad, E., & Spöttl, C. (2017). *Looking into listening: Using eye-tracking to establish the cognitive validity of the Aptis listening test*. ARAGs Research Reports Online AR-G/2017/3. British Council. Retrieved from https://www.britishcouncil.org/sites/default/files/looking_into_listening.pdf

Huckin, T. & Coady, J. (1999) Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *21*, 181–193. http://dx.doi.org/10.1017/S0272263199002028

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.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Elsevier Science.

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, *11*, 149–158. https://doi.org/10.1177/026765839501100206

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?. *Applied Linguistics*, *19*, 357–377. https://doi.org/10.1093/applin/19.3.357

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *18*, 178–191. https://doi.org/10.1037/a0026923

Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*, 755–780. https://doi.org/10.1017/S0272263118000165

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kasper, G. (1998). Analysing verbal protocols. *TESOL Quarterly*, *32*, 358–362. https://doi.org/10.2307/3587591

Kottmeyer, A. M., Van Meter, P., & Cameron, C. (2020). Diagram comprehension ability of college students in an introductory biology course. *Advances in Physiology Education*, *44*, 169–180. https://doi.org/10.1152/advan.00146.2018

Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied Psycholinguistics*, *40*, 1103–1139. https://doi.org/10.1017/S0142716419000158

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Krashen, S.D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45–77). Academic Press.

Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. Longman.

Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *International Review of Applied Linguistics in Language Teaching*, *28*, 293– 307. https://doi.org/10.1515/iral.1990.28.4.293

Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140–155). Cambridge University Press.

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. http://doi.org/10.3138/cmlr.59.4.567

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436. http://doi.org/10.1111/j.0023-8333.2004.00260.x

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, *42*, 625-651. https://doi.org/10.1017/S0272263120000108

Leow, R.P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.

Lessard-Clouston, M. (2009). Definitions in theology lectures: Implications for vocabulary learning. *The Asian ESP Journal*, *5*, 7–22. https://digitalcommons.biola.edu/faculty-articles/221

Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press

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

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207. https://doi.org/10.1111/lang.12117

Loewen, S. (2015). *Introduction to instructed second language acquisition*. Routledge.

Long, M. (1985). Input and second language acquisition theory. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Newbury House.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), The *handbook of second language acquisition* (pp. 413–468). Academic Press.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education*, *14*, 225–244. https://doi.org/10.1007/BF03172967

Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction*, *13*, 157–176. https://doi.org/10.1016/S0959-4752(02)00018-X

Lüdecke D. (2022). *sjPlot: Data visualization for statistics in social science*. R package (Version 2.8.11). Retrieved from https://CRAN.R-project.org/package=sjPlot

Lüdecke D., Ben-Shachar M., Patil I., Waggoner P., Makowski D. (2021). “performance: An R Package for Assessment, Comparison and Testing of Statistical Models.” *Journal of Open Source Software*, *6*, 3139. https://doi.org/10.21105/joss.03139

Lynch, T., & Mendelsohn, D., (2010) Listening. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd ed., pp. 180–196). Routledge.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, *43*, 985–1008. https://doi.org/10.1017/s0272263121000036

Malicka, A., Gilabert Guerrero, R., & Norris, J. M. (2019). From needs analysis to task design: Insights from an English for specific purposes context. *Language Teaching Research*, *23*, 78-106. https://doi.org/10.1177/1362168817714278

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mayer, R. (2014a). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–368). Cambridge University Press.

Mayer, R. (2014b). Introduction to Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 1–24). Cambridge University Press.

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

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: role of speaker's voice. *Journal of Educational Psychology*, *95*, 419–425. https://doi.org/10.1037/0022-0663.95.2.419

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. ED No. 04-CO-0040, U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. Retrieved from https://repository.alt.ac.uk/629/1/US_DepEdu_Final_report_2009.pdf

Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL*. Macmillan.

Mohamed, A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies In Second Language Acquisition*, *40*, 269–293. https://doi.org/10.1017/s0272263117000092

Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: An eye-tracking study. *The Language Learning Journal*, *47*, 460–478. https://doi.org/10.1080/09571736.2019.1638623

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*, 308–328. https://doi.org/10.1111/modl.12215

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

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

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts (Version 3) [Computer software]. Retrieved from https://www.lextutor.ca/cgi-bin/range/texts/index.pl

Nguyen, C. D., & Boers, F. (2018). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly*, *53*, 5–29. https://doi.org/10.1002/tesq.441

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, *24*, 517–537.

https://doi.org/10.1177/0265532207080771

O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, *10*, 418–437. https://doi.org/10.1093/applin/10.4.418

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.

Paivio, A. (2014). Bilingual dual coding theory and memory. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of bilingual memory* (pp. 41–62). Springer.

Paribakht, T. & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, *11*, 68–83. https://doi.org/10.18806/tesl.v11i1.623

Parry, K. (1991). Building a vocabulary through academic reading. *TESOL Quarterly, 25*, 629–652. https://doi.org/10.2307/3587080

Patanasorn, C. (2010). *Effects of procedural content and task repetition on accuracy and fluency in an EFL context* [Unpublished doctoral dissertation]. Northern Arizona University. Retrieved from https://www.proquest.com/dissertations-theses/effects-procedural-content-task-repetition-on/docview/759842793/se-2

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

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

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, A.

(2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, *42*, 577–598. https://doi.org/10.1017/S0272263120000091

Prabhu, H. (1987). *Second language pedagogy*. Oxford University Press.

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.

QSR International. (2018) NVivo (Version 12) [Computer software]. Retrieved from https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355–371. https://doi.org/10.1177/026553229301000308

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41–60). Erlbaum.

Révész, A. (2021). Exploring task-based cognitive processes: Methodological advances and challenges. *TASK*, *1*, 266–288. https://doi.org/10.1075/task.21017.rev

Révész, A., & Shi, D. (in prep). Learning through listening. In E. Wagner, A. Batty & Galaczi, E. (Eds.), The *Routledge handbook of second language acquisition and listening*. Routledge.

Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning*, *45*, 283–331. https://doi.org/10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge University Press.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, *22*, 27-57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2003). Attention and memory during SLA. In C.J. Doughty & M.H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678*)*. Blackwell.

Rost, M. (2011). *Teaching and researching listening*. Longman.

Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, *37*, 31–53. https://doi.org/10.1177/0265532219871470

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*, 759–769. https://doi.org/10.3758/BF03202637

Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In C. Was, F. Sansost, & B. Morris (Eds.), *Eye-tracking technology applications in educational research* (pp. 143–165). Information Science Reference.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). University of Hawai‘i.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge University Press.

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide* (Version 2.0) [Computer software]. Psychology Software Tools. Retrieved from https://support.pstnet.com

Schnotz, W. (2014). Integrated Model of Text and Picture Comprehension. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 72–103). Cambridge University Press.

Schnotz, W., & Wagner, I. (2018). Construction and elaboration of mental models through strategic conjoint processing of text and pictures. *Journal of Educational Psychology*, *110*, 850–863. https://doi.org/10.1037/edu0000246

Seaman, J. E., Allen, I. E., & Seaman, J. (2018). *Grade Increase: Tracking Distance Education in the United States*. Babson Survey Research Group. https://files.eric.ed.gov/fulltext/ED580852.pdf

Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text?. *TESOL Quarterly*, *52*, 971–994. https://doi.org/10.1002/tesq.445

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*, 165–179. https://doi.org/10.1017/S0272263100011943

Shintani, N. (2012a). Repeating input-based tasks with young beginner learners. *RELC Journal*, *43*, 39–51. https://doi.org/10.1177/0033688212439322

Shintani, N. (2012b). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research*, *16*, 253–279. https://doi.org/10.1177/1362168811431378

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

Skehan, P. & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp.183–205). Cambridge University Press.

Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning*, *17*, 517–556. https://doi.org/10.1080/0958822042000319692

Sonbul, S., & Schmitt, N. (2012). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*, 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

SR Research. (2016). Eyelink 1000 Plus brochure [Brochure]. Retrieved from https://www.sr-research.com/wp-content/uploads/2018/01/EyeLink-1000-Plus-Brochure.pdf

Staub A. & Rayner K. (2007) Eye movements and online comprehension processes. In Gaskell M.G. (Ed.) *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford University Press.

Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing*, *32*, 463–483. https://doi.org/10.1177/0265532214562099

Suvorov, R. (2018). Test takers' use of visual information in an L2 video-mediated listening test: Evidence from cued retrospective reporting. In E. Wagner & G. J. Ockey (Eds.), *Assessing L2 listening: Moving towards authenticity* (pp. 145–160).

John Benjamins.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury.

Swain, M. (2005). The output hypothesis: theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Routledge.

Tanenhaus, M. K. (2007). Spoken language comprehension: insights from eye movements. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 209–329). Oxford University Press.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. Traxler, M. Gernsbacher, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 863–900). Elsevier Science & Technology.

Tomlin, R.S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183–203. https://doi.org/10.1017/S0272263100012870

Tremblay, A., & Ransijn, J. (2020). LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models. R package (Version 3.0). Retrieved from https://cran.r-project.org/web/packages/LMERConvenienceFunctions/LMERConvenienceFunctions.pdf

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge University Press.

van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, *25*, 325–331. https://doi.org/10.1016/j.chb.2008.12.021

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*, 609–624. https://doi.org/10.1016/j.system.2013.07.012

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*, 191–210. https://doi.org/10.1017/S0261444807004338

Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research.* Ablex.

VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.) *Processing instruction: Theory, research, and commentary* (pp. 5–31). Routledge.

VanPatten, B., Williams, J., & Rott, S. (2004). *Form-meaning connections in second language acquisition*. Lawrence Erlbaum Associates.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition?. *Applied Linguistics*, *24*, 56–89. https://doi.org/10.1093/applin/24.1.56

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology*, *11*, 67–86. https://www.lltjournal.org/item/2604

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, *27*, 493–513. https://doi.org/10.1177/0265532209355668

Wagner, E. & Ockey, G. J. (2018). An overview of the use of audio-visual texts on L2 listening tests. In G. J. Ockey & E. Wagner (Eds), *Assessing L2 listening. Moving towards authenticity* (pp. 129–144). John Benjamins.

Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning*. https://doi.org/10.1111/lang.12495

Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics*, *28*, 46–65. https://doi.org/10.1093/applin/aml048

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

Wesche, M., Paribakht, S., & Haastrup, K. (2010). Research on the lexical inferencing process and its outcomes. In M. Wesche & S. Paribakht (Eds.), *Lexical inferencing in a first and second language: Cross-linguistic dimensions* (pp. 3–31). Multilingual Matters.

Williams, J.N. (2009). Implicit learning in second language acquisition. In W. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–344). Emerald.

Winke, P., & Gass, S. (2016). Using free recall and idea units for evaluating second language comprehension: Methodological choices and issues. In *TESOL Applied Linguistics Forum*. Retrieved from http://newsmanager.commpartners.com/tesolalis/issues/2016-11-04/5.html

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning and Technology*, *14*, 65– 86. http://dx.doi.org/10125/44203

Winke, P., & Lim, H. (2014). *The effects of testwiseness and test-taking anxiety on L2 listening test performance: A visual (eye-tracking) and attentional investigation*. IELTS Research Report Series. British Council/IELTS Australia. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_online_rr_2014-3.ashx

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

Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition*, *21*, 243–258. https://doi.org/10.1017/S0272263199002053

Wolvin, A., & Coakley, C. G. (1996). *Listening* (5th ed.). Brown & Benchmark.

Yanagisawa, A., & Webb, S. (2019). Measuring depth of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 371–386). London, UK: Routledge.

Yang, H. C., & Sun, Y. C. (2013). It is more than knowledge seeking: Examining the effects of OpenCourseWare lectures on vocabulary acquisition in English as a foreign language (EFL) context. *Computer Assisted Language Learning*, *26*, 1–20. https://doi.org/10.1080/09588221.2011.624523

Appendices

Appendix 1 Original Information Sheet

Information Sheet for Advanced Learners of English

09/2019-09/2022

My name is Danni Shi, and I am a doctoral student in the Department of Communication, Culture, and Media at the UCL Institute of Education, University College London. My research interests lie in the field of second language acquisition (SLA), and I am writing to invite you to participate in my research project. This project has been reviewed and approved by the UCL IOE Research Ethics Committee. Its primary objective is to investigate cognitive processes involved in comprehending multimodal lectures. I sincerely hope that you would be interested in participating.

1. Who is carrying out the research?
The researcher herself will be carrying out the research.

2. Why are we doing this research?
The results of this research will have implications for language teachers and second language learners, as they will contribute to our understanding of the essential abilities needed for successful academic listening.

3. Why am I being invited to take part?
I am looking for participants whose first language is Mandarin and who are at a higher intermediate to advanced proficiency level.

4. What will happen if I choose to take part?
If you decide to participate, I will first ask you to take an English proficiency test, a vocabulary size test, and a background questionnaire. Based on your performance, you may or may not be invited to attend further sessions. If invited, you will be asked to complete a video-lecture-based task. The second experimental session will take place in an eye-tracking lab at the Institute of Education. During the process, your eye movements will be recorded by a camera. One week later, a third session will be held during which you will be asked to complete several tests measuring individual differences such as working memory capacity.

5. Will anyone know I have been involved?
Any data obtained from you will be kept securely. At every stage of the project and beyond, your name will remain confidential. Your identity will be anonymised by the use of a unique identifier.

6. Could there be problems for me if I take part?
This research has no potential risk. No sensitive data will be collected from you.

7. What will happen to the results of the research?
After the study is finished, I plan to publish the results and present them at academic conferences. Please note that all data presented will be kept anonymous. Moreover, I will provide you with a summary of the overall findings of my research upon completion.

8. Do I have to take part?
Participation in this study is voluntary, and it is entirely up to you to decide whether or not to take part. You are free to withdraw from the study at any time without providing a reason, and this will not have any impact on you. I hope that you will find the experience beneficial if you choose to participate.

9. Contact for further information
If you require additional information before making a decision about participating, please do not hesitate to contact me.

Thank you very much for taking the time to read this information sheet.

If you would like to participate in the project, please complete the following consent form.

Appendix 2 Original Consent Form

CONSENT FORM

If you are happy to participate in this study, please complete this consent form and return it to Danni Shi in person or at the address below.

		YES	NO
1.	I have read and had explained to me the Information Sheet relating to this project.	□	□
2.	I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements for my participation as described in the Information Sheet.	□	□
3.	I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.	□	□
4.	I agree with the contents of this Consent Form and have received the accompanying Information Sheet.	□	□

Name: Signed:

Date:

Danni Shi

Department of Communication, Culture and Media

UCL Institute of Education

20 Bedford Way London WC1H 0AL

Appendix 3 Revised Information Sheet

Note: The modifications made to the information sheet have been highlighted for clarity.

Information Sheet for Advanced Learners of English

09/2019-09/2022

My name is Danni Shi, and I am a doctoral student in the Department of Communication, Culture, and Media at the UCL Institute of Education, University College London. My research interests lie in the field of second language acquisition (SLA), and I am writing to invite you to participate in my research project. This project has been reviewed and approved by the UCL IOE Research Ethics Committee. Its primary objective is to investigate cognitive processes involved in comprehending multimodal lectures. I sincerely hope that you would be interested in participating.

1. Who is carrying out the research?
The researcher herself will be carrying out the research.

2. Why are we doing this research?
The results of this research will have implications for language teachers and second language learners, as they will contribute to our understanding of the essential abilities needed for successful academic listening.

3. Why am I being invited to take part?
I am looking for participants whose first language is Mandarin and who are at a higher intermediate to advanced proficiency level.

4. What will happen if I choose to take part?
If you decide to participate, I will first ask you to take an English proficiency test, a vocabulary size test, and a background questionnaire. Based on your performance, you may or may not be invited to attend further sessions. If invited, you will be asked to complete a video-lecture-based task. The second experimental session will take place in an eye-tracking lab at the Institute of Education. During the process, your eye movements will be recorded by a camera. Please read Appendix 1 for more details about the eye-tracking session. One week later, a third session will be held during which you will be asked to complete several tests measuring individual differences such as working memory capacity.

5. Will anyone know I have been involved?
Any data obtained from you will be kept securely. At every stage of the project and beyond, your name will remain confidential. Your identity will be anonymised by the use of a unique identifier.

6. Are there any risks for me if I take part?
There are no potential risks associated with the experimental sessions, and no sensitive data will be collected from you. Regarding COVID-19 risks, strict guidelines that have been put in place by UCL and the government will be followed. To minimise exposure to COVID-19 for participants, except for the eye-tracking session, other tests and questionnaires will be administered online. Eye-tracking data will be collected from one participant at a time in the language lab. Participants and the researcher will need to wear masks and maintain social distancing throughout the session. Items used as part of the study (e.g., keyboard, computer mouse, headphone, etc.) will all be thoroughly sanitised before and after being used. However, a certain level of risk is unavoidable. For more information on the additional COVID-19 procedures, see Appendix 2.

7. What will happen to the results of the research?
After the study is finished, I plan to publish the results and present them at academic conferences. Please note that all data presented will be kept anonymous. Moreover, I will provide you with a summary of the overall findings of my research upon completion.

8. Do I have to take part?
Participation in this study is voluntary, and it is entirely up to you to decide whether or not to take part. You are free to withdraw from the study at any time without providing a reason, and this will not have any impact on you. I hope that you will find the experience beneficial if you choose to participate.

9. Who do I speak to if problems arise?
If you have any complaints about the way in which this research project has been, or is being, conducted, please in the first instance discuss them with the researcher. If the problems are not resolved, or you wish to make a formal complaint, please contact Professor Andrea Révész, who is the principal supervisor of the researcher.

10. Contact for further information
If you require additional information before making a decision about participating, please do not hesitate to contact me.

Thank you very much for taking the time to read this information sheet.

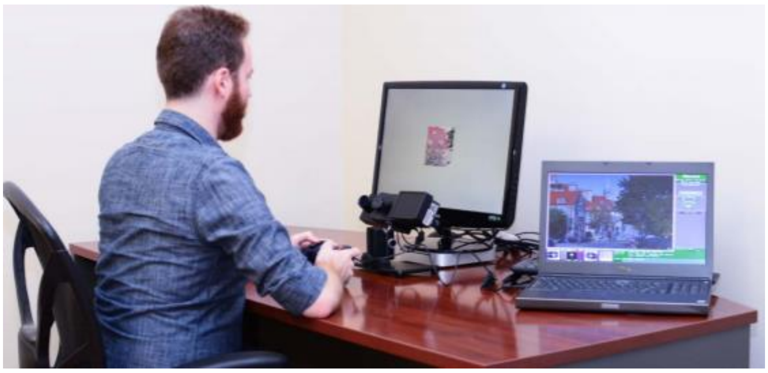
If you would like to participate in the project, please complete the following consent form.

Appendix 1 Details about the Eye-Tracking Session

Eye tracking is a technique that measures where a person is looking and how long they focus on a particular spot. It is often used to investigate learners' cognitive processes, given that eye movements can reflect what we pay attention to and the cognitive effort required to process information.

Before each eye-tracking session, I will introduce the eye-tracking device used in my study and explain the procedure to you in detail. During the session, you will be seated in front of a computer screen and then complete a calibration check by looking at a point that appears on the screen. Following that, you will watch a two-minute example video, after which you will complete another calibration check.

Next, you will watch an academic lecture while the eye tracker records your eye movements. I will use the Remote Mode of the eye tracker, which allows for natural head movements. However, excessive movement may affect the quality of the eye-tracking data, so I will monitor your eye movements from a separate room during the task. The eye-tracking session will last approximately 30 to 90 minutes.



Appendix 2 Information for participants concerning additional COVID-19 procedures

This document outlines the additional measures I have taken to minimise the risk of COVID-19 transmission during your visit to the eye-tracking lab at the IOE. The procedures described are designed to minimise the known risks of COVID-19 infection, but cannot abolish all risks. Please review this information carefully and do not hesitate to contact the researchers if you have any questions or require clarification before deciding whether to participate in the study.

Before your appointment:

It is very important that you read the information sheet and consent form in relation to the study carefully prior to your participation. If you have any questions about completing any part of the consent form or require clarification before deciding whether to participate in the study, the researcher will arrange an online meeting to answer any questions you may have. Next, I will get in touch with you 48 hours before your scheduled eye-tracking session to review the specifics of your visit. The researcher and all participants will need to take rapid COVID-19 tests up to 24 hours before the eye-tracking session. This is put in place to ensure the safety and well-being of all involved.

Arrival at the IOE:

Please ensure that you arrive on time, but no earlier than 10 minutes prior. This is important to minimise contact with other students in the building and to avoid crowding in communal areas. Upon your arrival, I will meet you at the main entrance located on the east side (Bedford Way). Please kindly note that no food or drinks are allowed in the lab.

During the eye-tracking session:

- During the visit to the eye-tracking lab, all participants must wear a face mask throughout the whole visit. The researcher will be wearing personal protective equipment, such as plastic visors covering the face and plastic gloves.
- All participants will be required to sanitise their hands regularly as instructed by the researcher.
- Upon arrival, temperatures will be taken and a questionnaire will be administered to all participants asking questions about COVID-related symptoms.
- Only two persons will be present in the lab at a time, including the researcher and one participant. Social distancing must be maintained at all times.
- Strict disinfection procedures will be followed before and after experimental sessions.

Potential risks:

The procedures implemented are intended to mitigate the risk of COVID-19 infection as much as possible. However, a certain level of risk is unavoidable. I recommend that you consider walking, cycling, or driving to the Institute of Education. If you choose to use public transport, please be aware of the risk of COVID-19 exposure through contact with other passengers. Additionally, you must adhere to government guidelines while using public transport.

Appendix 4 Revised Consent Form

Note: The modifications made to the consent form have been highlighted for clarity.

CONSENT FORM

If you are happy to participate in this study, please complete this consent form.

	YES	NO
1. I have read and had explained to me the Information Sheet relating to this project.	□	□
2. I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements for my participation as described in the Information Sheet.	□	□
3. I understand that my participation is entirely voluntary and that I have the right to withdraw from the project at any time.	□	□
4. I have read and understand the information for procedures in place for reducing COVID-19 transmission risks. I have had an opportunity to consider the information and what will be expected of me.	□	□
5. I agree with the contents of this Consent Form and have received the accompanying Information Sheet.	□	□

Name: Signature:

Date:

Danni Shi

Department of Communication, Culture and Media

UCL Institute of Education

20 Bedford Way London WC1H 0AL

Appendix 5 UCL Institute of Education Risk Assessment Form

UCL Institute of Education Risk Assessment Form

Starting or Resuming Fieldwork in the contexts of COVID-19

This risk assessment form relates to the IOE Starting or Resuming Face-to-Face Fieldwork - Response to Research and changed contexts due to COVID-19. Before completing this form, please make sure you have read this guidance and are familiar with the process for approving projects that aim to start or resume fieldwork.

This form has been designed so that UCL IOE researchers can:

- be made aware of the risks and hazards involved in starting or resuming fieldwork and face-to-face data collection during the global COVID-19 pandemic, and what steps to take to mitigate them
- indicate the risks and hazards they may encounter in their fieldwork for their specific research project
- indicate what measures they plan to put in place to mitigate these risks and hazards

Actions required – please:

1. **Read** the list of risks, hazards and control measures in the Appendix
2. **Complete** all sections of this form in full, providing as much detail as possible and referring to the guidance in the Appendix
3. **Email** this form to ioe.researchethics@ucl.ac.uk who will then formally submit it to Prof Phil Jones (IOE REC Chair) & Simon Buller (Director of Operations) for approval

---

Section 1: Project Details

a. Name(s) of researcher(s) submitting this risk assessment: Danni Shi

b. Name of supervisor (if applicable): Andrea Révész

c. Research project title: The effects of task repetition on learners’ processing of multimodal L2 input and acquisition of technical vocabulary

d. UCL IOE department: Culture, Communication and Media

e. Is this project:

A resumption of a project that has already been granted full ethical approval by the IOE Research Ethics Committee: ☒

- Project REC code: N/A

A new project that has not yet received full ethical approval: ☐

f. Brief description of project, including location(s) of fieldwork (maximum 300 words):

The aim of the current study is to investigate how task repetition affects second language (L2) learners' multimodal processing and vocabulary acquisition from viewing. Participants will be 75 postgraduate students at the UCL Institute of Education (IOE). The participants will be invited to take part in three sessions. The first and third sessions will be carried out online, while the second session will be held in an eye-tracking lab at the IOE. In the second session, the participants will be asked to watch a video lecture once, twice, or three times, while their eye movements will be recorded during each viewing. Eye-tracking data will be collected from one participant at a time.

Section 2: Project Participants

Please indicate below the categories of people potentially at risk from any activities within your research project and that will be covered by this risk assessment:

Researcher(s): ☒
Members of the public: ☐
Research participants: ☒
Other: ☐

More details:

The researcher and participants may be exposed to COVID-19 while travelling to and from the IOE and also during the eye-tracking sessions.

Section 3: Risk Assessment

a. Please read the **Appendix** on **COVID-19 risks and hazards** (pg. 3 onwards), and enter details of the potential risks arising from the work planned for this project below.

For example, '*members of research team will have to use public transport to travel to the school where fieldwork will take place'*

1. Participants who use public transport or a taxi to travel to and from the IOE may be at risk of contracting COVID-19 if they come into contact with infected individuals or touch contaminated surfaces.

2. During the eye-tracking session, there is a risk of COVID-19 infection for the researcher and participants due to the possibility of contact with infected individuals or contaminated surfaces.

b. Please confirm the steps you will take to mitigate the above COVID-19-related hazards and risks. See the Appendix for some examples of control measures. Please reference specific control measures or sections of the Appendix, where possible.

For example, (in reference to the above) *'members of research team will be asked to plan their journeys ahead to avoid busy times. They will be asked to wear a face covering while travelling, as well as trying to socially distance as much as possible.*

They will be encouraged to wash their hands or use hand sanitiser once their journey is complete'

1. Before the eye-tracking session, the following guidelines will be adhered to:

(a) The researcher will contact the participants 48 hours before an eye-tracking session to review the specifics of their visit and check their state of health. Additionally, the researcher and participants will take symptom-free Covid-19 lateral flow tests (LFT) up to 24 hours before an eye-tracking session. The session will continue only if the results are negative.

(a) All eye-tracking sessions will be scheduled at an off-peak time so that the participants can avoid using public transport during peak hours. The researcher will walk to the IOE to minimise the risk of exposure to COVID-19. The participants will be encouraged to walk, cycle, or drive to the IOE. If they have to use public transport, the participants will be asked to ensure that they cover their face with a mask while travelling, as well as trying to socially distance themselves from other passengers as much as possible. Once they complete their journey and arrive at the IOE, they will then be given a fresh, disposable mask to wear. Then, they will be taken to a washroom near the eye-tracking lab to wash their hands and use hand sanitiser.

2. For all eye-tracking sessions, the following guidelines will be adhered to:

(a) Temperatures will be taken prior to an eye-tracking session. Questionnaires asking questions about COVID-related symptoms will be administered as well.

(b) There will be no more than 2 persons in the eye-tracking lab, including the researcher and one participant.

(c) Both the researcher and the participant need to wear masks throughout the session. The researcher will also wear additional protection equipment (i.e., plastic visors covering the face and plastic gloves).

(d) Social distancing between the researcher and the participant will be maintained throughout the session. The eye-tracking system used in the study consists of a desk-mounted camera, a display computer, and a host laptop. The participant will be instructed to watch a video lecture using the display computer in a room, while the host laptop will be put in a separate room for the researcher to monitor the participant's eye movements throughout the task.

(e) Items used as part of the study (e.g., keyboard, mouse, headphone, etc.) will all be thoroughly sanitised before and after being used.

Section 4: Declaration

Researcher's signature/name: Danni Shi

Date: 08/03/2021

Appendix 6 IOE Postgraduate Research (PGR) Student Resuming Fieldwork Proforma

Notice to request starting face-to-face fieldwork in the contexts of COVID-19

This form is to be completed for all IOE research projects that intend to start face-to-face fieldwork during the COVID-19 pandemic. This form enables researchers and our ethics reviewers to focus on issues specifically concerning fieldwork in the contexts of COVID-19.

As part of the process to request starting fieldwork, the Principal Investigators, or a nominated Co-Investigator or Research Assistant/Associate, will need to submit to Moodle:

- An ethics application form with attachments such as information sheets and consent forms
- This form, fully completed
- An IOE Fieldwork Risk Assessment form

Before completing this notice please make sure you have read the 'IOE starting or resuming face-to-face fieldwork' guidance.

If you are reviewing this form, please answer all questions in the yellow highlighted 'reviewer feedback' sections, then email this form back to ioe.researchethics@ucl.ac.uk

Section 1 – Project Details

a. Project Title The effects of task repetition on learners' processing of multimodal L2 input and acquisition of technical vocabulary
b. Principal Investigator (PI) Danni Shi

c. Please provide a summary of face-to-face research activities below:

The aim of the current study is to investigate how task repetition affects second language (L2) learners' multimodal processing and vocabulary acquisition from lecture viewing. Participants will be 75 postgraduate students at the UCL Institute of Education (IOE). The participants will be invited to take part in three sessions. The first and third sessions will be carried out online, while the second session will be held in an eye-tracking lab at the IOE. In the second session, the participants will be asked to watch a video lecture once, twice, or three times, and their eye movements will be recorded during each viewing. Eye-tracking data will be collected from one participant at a time.

Section 2 – Rationale for Starting Fieldwork

The UCL guidance states *'the researcher should verify whether the fieldwork can be carried out remotely before making any plans for person-to-person interaction'*.

Please clarify below how you have verified that data could not be collected remotely for this project. In particular, how have you addressed the ethical obligation to not place extra burden on participants, such that risks outweigh benefits for any participant?

My doctoral study involves collecting eye-movement data from second L2 learners to investigate their cognitive processes underlying multimodal lecture comprehension and vocabulary acquisition. The study utilized an eye-tracking system, including a display computer, an eye tracker, and a host laptop, that could not be remotely administered. To minimise the risk of COVID-19 infection, except for eye-tracking sessions, other experimental sessions will be conducted remotely via Zoom meetings. Eye-tracking sessions will be carried out strictly following UCL and government guidelines on social distancing, hand washing, and respiratory hygiene. Although this research project may not bring any immediate benefits to the participants, they will have the opportunity to engage in a reflective activity which is often included in professional development workshops. In addition, the information obtained from the project will help inform the development of more efficient and effective L2 learning methods. This is especially important in the COVID-19 context, as video-based lectures have become a major component of university education.

Section 2 - Reviewer feedback:

Has the P.I. provided a satisfactory justification for the commencement of fieldwork?

Click or tap here to enter text.

Section 3 – Organisation/Site Policies

3a. In relation to the site of the intended fieldwork, have you attached a copy of the organisation/site policies (or equivalent) in respect of COVID-19?

☒ - Yes ☐ - No

3b. Have you obtained agreement with the organisation/site that they approve of the arrangements made for the fieldwork to start?

☐ - Yes ☒ - No

If no, please describe your process for obtaining an agreement with the organisation/site, and confirm that this will be sent to the IOE Research Ethics Office once obtained.

The eye-tracking data will be collected in an eye-tracking lab at the IOE. To obtain an agreement with the IOE, I need to revise my original ethics application form, information sheet, and consent form, before sending these documents along with a rationale for the revisions, the IOE Fieldwork Risk Assessment, IOE staff starting fieldwork proforma, and related safety guidance to my supervisor for approval. Once the above is approved by my supervisor, I need to submit all documents to the Department Graduate Tutor for further processing. Then the IOE Research Ethics Committee (REC) will review my proposed face-to-face data collection application and make a decision. The approval for collecting face-to-face data will be sent to the IOE Research Ethics Office once obtained.

3c. Please detail below how the conduct of this research will comply with organisation/site policies in respect of COVID-19, and how compliance guarantees safety for participants and researchers.

1. Minimising infection transmission when travelling to and from fieldwork

− The researcher will walk to the IOE.

− All eye-tracking sessions will be scheduled at an off-peak time so that participants can avoid using public transport during peak hours. They will be encouraged to walk, cycle or drive to the IOE. If they have to use public transport, they will be asked to follow the social distancing roles, cover their face with masks throughout their journey, and sanitise their hands before and after travelling.

2. Minimising infection transmission when collecting face-to-face data

− The researcher will contact participants 48 hours before their eye-tracking session to check their state of health. Both the researcher and all participants will need to take rapid COVID-19 tests up to 24 hours before the session to confirm their health status.

− The participants will be instructed to arrive on time, and the researcher will meet them at the main entrance to the IOE. This is to minimise interaction with other students in the building and to avoid crowding in communal areas.

− The participant will be given a fresh, disposable mask to wear, before being taken to a washroom near the eye-tracking lab to wash their hands and use hand sanitiser.

− There will be no more than 2 persons in the eye-tracking lab, including the researcher and one participant. Social distancing between the researcher and the participant will be maintained throughout the session. Both the researcher and the participant need to wear masks and sanitise hands regularly throughout the session.

– Items used as part of the study (e.g., keyboard, mouse, headphone, etc.) will all be thoroughly sanitised before and after being used.

3d. If the activities of this research raise issues that are not adequately or clearly covered by the existing organisation/site policies, please clarify how this will be addressed below.

The current study will strictly follow government guidelines if issues that are not clearly covered by the existing site policies raise. The researcher will also discuss the issues with her principal supervisor and contact the IOE REC for advice when necessary.

3e. Please detail below how any resources necessary to comply with organisation/site policies in respect of COVID-19 are in place.

A non-contact electronic forehead infrared thermometer will be used to measure body temperature the participants arrive at the IOE, followed by an electronic questionnaire asking questions about COVID-related symptoms. Protection equipment will be provided to all participants, including fresh disposable masks and hand sanitiser. Plastic visors covering the face and plastic gloves will be provided if requested. In addition, all participants and the researcher will have symptom-free COVID-19 lateral flow tests (LFT) up to 24 hours before an eye-tracking session. The participants can get access to the test at the UCL testing centre or local community LFT testing centres. The eye-tracking session will be carried out only if the test results are negative.

3f. Please make clear how you intend to address any issues concerning equality, diversity, and inclusion, and additional resources or adapted ways of working.
(Example of this are participants or researchers who are Deaf/deaf, or Neurodivergent people and those with cognitive impairment)

The participants will be postgraduate students at the IOE aged between 20 to 40. COVID-19 may have less of an impact on this age group as there is evidence showing that older age groups (particularly those over 50 years old) may be more at risk. The risks of participating in the study were included in the revised information sheet and will be explained to the participants in advance. Online meetings will be scheduled as well to answer questions and discuss concerns that participants may have. As for individuals that identified as more clinically vulnerable to COVID-19 (e.g., older participants, pregnant individuals, and people with disabilities), they will be asked to consider their participation carefully.

Section 3 – Reviewer feedback

Has the P.I. sufficiently addressed the above questions? If the P.I. has not, or only partially addressed the above, please elaborate either below, or under any above items that you feel need further clarity.

Click or tap here to enter text.

Having read the organisation/site's policies and the application's response, are there areas that you think need addressing in relation to fieldwork during COVID-19 that are not yet addressed or covered?

(Please draw on the UCL *'Working safely during the Pandemic'* webpage, which has updated sections on areas relevant to most situations connected to research)

Click or tap here to enter text.

Section 4 – Supporting Documents

4a. Have you provided information sheets for all participants and stakeholders that adequately reflect arrangements that address the contexts of COVID-19

☒ - Yes ☐ - No

Section 4 – Reviewer feedback

Has the P.I. provided appropriate information sheets for participants and stakeholders that adequately reflect these arrangements?

Click or tap here to enter text.

Appendix 7 Video Transcripts

Video 1: Parts of Neurons

So in this module we're going to meet the stars of the nervous system, and the stars of the nervous system are neurons. Neurons are really special. So, admittedly, I am a complete neuro-chauvinist. But, I can tell you that, let's say you've seen ten gastrointestinal cells, you've seen them all. Let's say you've seen ten cardiomyocytes, you've seen them all. Ten lung cells, liver cells, kidney cells, etc. Kind of pretty standard. There might be five, ten different types but after that you're, you're out of variation. Totally different story in the, in the department of neurons. So, neurons are, are basically unique. You could, if you were a real lumper, you could categorize them into maybe a few million different types. So, we are going to look at neurons in some detail. Before, before we get into their particular appearance, let's just understand one of the features that makes neurons so unique. And that is that they are the longest, they can be the longest cell in the body. So, our longest neuron is a neuron that has a process that starts at your big toe and ends right here. In me, that's about five feet, one and a half meters. In really tall people that can be six feet, two meters. So, that's a really long process. There's no other cell type in the entire body that is that long or that large. Now, in the next segment, what we're going to do is we're going to get the vocabulary to understand the, the individual uniqueness of of neurons.

There are four parts to neurons. Neurons have four parts. The first is the cell body, also called the Soma. And this is the part that all cells have, this is cell central. There's the nucleus is here with the DNA. This is like city hall, it's gonna give out all the orders. There's a manufacturing plant here where proteins are made. There is a power plant here. So this is really the place that keeps the cell going. It makes all the material for the entirety of a neuron which it can be, as we said in the last segment, very, very long. Very large. So the cell body has coming out from a number of dendrites, and these dendrites branch. And they continue to branch. And they can branch two, three, four, five times. And so, that makes a tree which we call the dendritic arbor or the dendritic tree. And these dendrites are responsible for gathering in information. They're the sentries. They are the ears of the cell, of the neuron. They're taking in all information. So information is going in to the dendrites, into the cell body to a certain extent, but really in to the dendrites. So what do we do with all that information? Well, the cell is going to process it, sum it up, and then send it out through one axon. There's one axon. And while these dendrites are very local, they're gonna have a local distribution. This axon can go far, far distances. And so this can go a meter, easily. So this is an axon, and it's gonna carry the information along the length of it. But then what? So you have to get to a terminal, which is a communication center. Which is gonna take that information and give it to the next cell in line. And if we just blow up this area right here. What we see is that there's a synaptic terminal, which is separated by a space. And then contacts another cell. So the information transfer is in this direction, from the axon to the next cell. And this point of information transfer is called the Synapse. This cell, this second cell, the cell that the neuron is talking to can be any of a variety of cell types. It can be another neuron. So most of the neurons in the brain, they just talk to other neurons. There's a big internal conversation going on there. So neurons talk to neurons. But neurons also talk to muscles. And neurons talk to glands. And neurons talk to the cardiac, to the heart, to the cardiac muscle and so on. Okay. In the next segment, what we're gonna do is look at the variety. I've drawn one cell, but cells look really different. Even in the same place, they look really different. And we'll look at some of the variety of neurons and talk a little bit more about their uniqueness.

Video 2: Meninges

So, we're going to talk now about the difference between the Central Nervous System and the Peripheral Nervous System. There's a barrier, there's a fence, and that fence is the demarcation between central and peripheral. And that fence is made up of three membranes, and those three membranes are the meninges. There are three meningeal layers, and the three layers go from very weak, very tender, the pia, to very tough, the dura and, in between, there's a spidery thing called the arachnoid, a spidery layer called the arachnoid. So, let's just look at what that looks like. We're only going to see the arachnoid in, in this next picture. So here, this is the brain, and what's covering the brain here, up here is pia and you can't see it, it's so thin, and so tender that we don't see it. But in this view, the dura has been removed and now we're looking down on this film, it's a, it's an actual membrane, and that membrane is the arachnoid. So, the different meningeal layers have different purposes, different functions. The dura is the, is a tough sack, and that is what is going to keep us from having concussions all the time. All right, so it's going to float our brain in in fluid and prevent it from banging about, and, and actually getting bruised. It takes a hard hit to actually injure the brain, and hopefully we, we don't have, we don't incur that. Okay, so the meninges go from, pia is closest to the central nervous system, is closest to the brain and spinal cord, and then dura is farthest away. And their, all the neurons, almost all of the neurons in the central nervous system, 200 billion minus something less than 100,000 are contained completely within the central neurons. The only neurons that leave the central nervous system, are these neurons that, that serve a motor function. They actually go out the meninges and they go into the periphery. In the periphery, there's the peripheral nervous system, and there's also every, the rest of the body. So there are motor neurons, they go to skeletal muscle, the voluntary muscle, so that we can move our arms and our legs. And these are motor neurons. And then there are Motor neurons that go to Autonomic Neurons that control our glands, that controls the cardiac muscle, they control our gut, smooth muscle. So for instance, these are going to be, these neurons are what is going to be active when you see food and you start to salivate. Well, somebody's got to tell that gland to start salivating and it's the brain and it does it through this two-neuron chain. So, those are Autonomic Neurons. In addition to the motor neurons that carry information from the central nervous system to the peripheral nervous system, there are sensory neurons. And these sensory neurons are peripheral. They are located peripherally, and they carry information into the central nervous system. What you'll note is we call a neuron either or central or peripheral based on where the cell body is not based on where its axon is or its dendrites, but based on where the cell body is. So these Peripheral Neurons include Sensory Neurons and Autonomic Neurons, these are in the autonomic ganglia. The consequence of that is that these neurons share vulnerabilities. The, the, they share with each other but not with the central neurons. So, for example, there's a disease called Congenital insensitivity to pain. People with this disease suffered a mutation in one of a number of developmental, developmentally important genes. And what it did was it, it prevented a group of sensory neurons from developing, those sensory neurons that respond to injury. So, these people are insensitive, they never feel pain, because they can't, they don't have the sensory neurons that will respond to noxious stimulation which is what usually brings us pain. But in addition, there's a group of the Autonomic Neurons that also developmentally got knocked out, and these are the neurons that lead you to be able to sweat. And so, people with Congenital insensitivity to pain often have what is called anhydrous, meaning they can't sweat. Those seem like two completely weird symptoms to put together, but they're not weird because they make complete sense. Both of them are due to a developmental abnormality in the peripheral nervous system. So, in the next segment what we're going to do is, is look again, look a little bit more closely at other diseases that preferentially affect either the periphery or the central nervous system.

Video 3: Rods and Cones

Okay, so, now we're going to talk about photo receptors, and they come in two varieties. Before we talk about the two varieties, the rods and the cones, let's talk about when we use vision. We use vision over this incredibly wide range of brightness. We can see a single photon when we've been in the dark. It's a, let's say it's a new moon. There's no light out. It's pitch black, and if there was a single photon, we would be able to detect that. Very, very dark conditions, even a small moon can give you what is called scotopic conditions, where there's not very much light. And on the other hand, we still can see in bright sunlight, and that's called photopic. And in between is is called mesopic. So, how many, how, how, how big is this range? Well, it is many orders of magnitude, which means it's, they are log units about, I think it's about 12 log units from scotopic to photopic. And, and that means that you can go from pitch black to bright, bright sunshine, and you can still see. And how do we do that? Well, in part, we do it by assigning these dimmer conditions to a different set of photoreceptors than these brighter conditions. And there's an overlap. There's an overlap here somewhere in, in the mesopic range, where both the the rods, which are useful in dim light conditions and the cones, which are the only things offered of, in photopic condition. So, there's this range where they're both operating. And if we come over here, I've drawn out a model of a one rod and one cone. So, they look different. And just to orient you, this is the opposite orientation from what we just had on, on the screen. The light is coming up in the instance. The retinal pigment aphelion is up at the top. And this area here, this is the outer segment. And the outer segment of the rod and of the cone look a little different. This looks like a paddle. This looks like half half a tree, half a tree, half a Christmas tree, say. And along the the surface or within this outer segment are all of these molecules of rhodopsin. This is the, the molecule that's going to turn light into neural energy. And what's interesting about how it, that is done, is that, let's take the situation of a rod. This is the rod. In the situation of the rod, there, the rodopsin molecules, are actually internal. They're inside the cell. They're sitting on membranes that are inside the cell. And so, when light comes in here, we'll get, get a new color. So, light's going to come in here. It's going to activate this molecule, and this is actually a metabotropic receptor, which means it doesn't actually have a channel, doesn't have a pore. But what it does, is it changes an enzyme, which then goes, and has to travel to the edge of the cell, and open or close a cell an ion channel. So, as you may imagine, the amount of time that this takes is actually long. It takes much longer than it takes, for instance, just to change it, an ion channel directly. And that's one of the reasons why vision is a very slow process. It doesn't really matter, because in the natural world, our, what we look at doesn't change on less than say, two millisecond time scale. Things aren't changing at any time scale that this would this would affect. Okay, so, that's how the photoreceptors work, and what's different about the rod and the cone? Well, the rod is extremely sensitive to light. So, it can respond to a very low number of photons hitting here. The cone is less sensitive. The, each of these has a preferred wavelength that it likes, and as it turns out, there is one type of rod, but in humans there are three types of cones. So, there are three different types, and each of those types has a wavelength of light that is preferred that it responds maximally to. And what that means is that the cones can be used. They don't produce color, but they can be used by the brain to enable us to perceive color. So, the cones are necessary for perceiving color. And so, it, if you're only using rods, could you use that information to perceive color? And the answer is no. And so, consequently, under scotopic conditions, there's no, there are no we cannot perceive color.

Under photopic conditions, we see, see vibrant colors. Under mesopic conditions, we see muted colors.

Appendix 8 Immediate and Delayed Vocabulary Post-Tests

1. Form recognition test

If you think the following word was mentioned in the lecture, please press “Y”. If not, please press “N”. You do not necessarily need to know the meaning of the word.

Example: apple (Press “N”)

dura

allele

mesopic

glia

syncope

gyrus

meninges

arachnoid

ischemic

telomere

scotopic

amyloid

soma

myelin

axon

photopic

enteric

dendrite

prion

thalamic

pia

2. Meaning recall test

Please write down anything you know about the following words. It can be a translation, a definition, or an explanation.

Example:

apple: 苹果 or a round fruit with red skin and white firm flesh.

dendrite:

axon:

mesopic:

pia:

meninges:

dura:

photopic:

arachnoid:

synapse:

soma:

scotopic:

3. Meaning recognition test

Please choose the closest meaning for the following words by pressing corresponding keys (“A”, “B”, “C”, “D”, and “E”). If you do not know the meaning of the word, please select the option “I don’t know”.

Example: apple (Press “B”)

A. a long vegetable with dark green skin

B. a round fruit with shiny red skin

C. a long curved fruit with a thick yellow skin

D. a round vegetable with large green leaves

E. I don’t know

dura

A. a projection of a cell body responsible for transmitting information

B. the process by which knowledge and understanding is developed in the mind

C. a tough brain membrane

D. a soft brain membrane

E. I don’t know

mesopic

A. the supporting cells of the central nervous system

B. under low but not quite dark lighting conditions

C. under very light conditions

D. three layers of brain membranes

E. I don’t know

meninges

A. three layers of brain membranes

B. a brain membrane that looks like a spider web

C. under low but not quite dark lighting conditions

D. a mass of abnormal cells found in the brain

E. I don’t know

arachnoid

A. three layers of brain membranes

B. under very bright conditions

C. the maintaining of the visual gaze on a single location

D. a brain membrane that looks like a spider web

E. I don’t know

scotopic

A. under very dark conditions

B. under low but not quite dark lighting conditions

C. the cell body

D. quick and simultaneous movements of both eyes

E. I don’t know

synapse

A. the cell body

B. a brain membrane that looks like a spider web

C. the transparent layer which covers and protects the outer part of the eye

D. the point of information transfer

E. I don’t know

soma

A. the cell body

B. the point of information transfer

C. under very dark conditions

D. the bone structure that forms the head and protects the brain

E. I don’t know

axon

A. a tough brain membrane

B. the way you notice things

C. a projection of a cell body responsible for transmitting information

D. an extension of a cell body responsible for gathering in information

E. I don’t know

photopic

A. the point of information transfer

B. the clear, colorless liquid found surrounding the brain

C. under very bright conditions

D. under very dark conditions

E. I don’t know

dendrite

A. the outer layer of the brain

B. an extension of a cell body responsible for gathering in information

C. a projection of a cell body responsible for transmitting information

D. a soft brain membrane

E. I don’t know

pia

A. a sudden serious illness when a blood vessel in the brain bursts or is blocked

B. a soft brain membrane

C. a tough brain membrane

D. an extension of a cell body responsible for gathering in information

E. I don’t know

Appendix 9 Questionnaires

1. Background questionnaire

(1) Age: ______________________

(2) Gender: □ male □ female □ prefer not to say

(3) Current level of study:

□ Undergraduate Year of study: □ 1st □ 2nd □ 3rd Other: _____

□ Postgraduate (taught) Year of study: □ 1st □ 2nd Other: _____

□ Postgraduate (research) Year of study: □ 1st □ 2nd □ 3rd □ 4th Other: _____

(4) Major subject: _____________________

(5) Standardized English proficiency test score:

Test	Overall score	Listening	Reading	Speaking	Writing	Date of taking the test (mm/yy)
IELTS						
TOEFL						
Other (Please specify):						

(6) Months of residence in English speaking countries: __________________

(7) Months of residence in the UK: ___________________________________

(8) Months of formal education in English speaking countries: _____________

(9) How many years have you been learning English? _________________________

(10) Estimate your level of English on a scale of 1 (beginner) to 5 (advanced learner)

Listening	1	2	3	4	5	Speaking	1	2	3	4	5
Reading	1	2	3	4	5	Writing	1	2	3	4	5

(11) What is your first language? ______________________________

(12) Have you ever learned any other languages except English?

□ Yes Please specify the language and your proficiency level:

____

□ No

(13) Have you ever taken any online courses taught in English? Yes □ No □

(14) Have you ever taken any online *Coursera* courses? Yes □ No □

(15) Which English accent do you prefer?

□ British accent □ American accent

2. Perception questionnaire

For each of the following statements, please put a tick (✓) in the column that best represents your level of agreement.

Statement	Strongly disagree	disagree	Neural	agree	Strongly agree
1. The topic of the lecture was familiar.					
2. The task was easy.					
3. I did well on this task.					
4. I invested little mental effort to complete this task.					
5. The words in the lecture were easy.					
6. I understood the content of the lecture.					
7. I found it easy to take notes while listening to the lecture.					

3. Post-experiment questionnaire

For each of the following vocabulary items, please put a tick (✓) in the column that best indicates your knowledge of that word.

Item	I didn't know this word before watching the video	I don't think I knew this word before watching the video	I have no idea if I knew this word before watching the video	I think I knew this word before watching the video	I definitely knew this word before watching the video
photopic					
dendrite					
dura					
scotopic					
meninges					
pia					
axon					
arachnoid					
synapse					
mesopic					
soma					

Appendix 10 COVID-19 Pre-Experiment Questionnaire

Please complete the following COVID-19 Pre-experiment questionnaire.

1. Within the last 14 days, have you experienced a new cough that you cannot attribute to another health condition?

□ Yes □ No

2. Within the last 14 days, have you experienced new shortness of breath that you cannot attribute to another health condition?

□ Yes □ No

3. Within the last 14 days, have you experienced a new sore throat, loss of taste or smell that you cannot attribute to another health condition?

□ Yes □ No

4. Within the last 14 days, have you had a temperature at or above 37.8 °C or the sense of having a fever?

□ Yes □ No

5. Within the last 14 days, have you had close contact with someone who is or was sick with suspected or confirmed COVID-19? Please note that close contact is defined as within 6 feet for more than 10 consecutive minutes.

□ Yes □ No

6. Within the last 14 days, have you or a household member been isolating?

□ Yes □ No

Please return this questionnaire to the researcher. If you answer yes to any of these questions, your eye-tracking session will be rescheduled.

Appendix 11 Instructions for Stimulated Recall and Data Excerpts

1. Instructions for research participants (translated from Mandarin):

Let's take a look at the recording of your eye movements while you were watching the short video before the lecture. This little pink dot on the screen shows where you were looking, and we can also hear what the speaker was saying. But I'm more interested to know what is going on inside your head at that time. So, we will now watch the recordings of your eye movements during lecture viewing, and I hope you could share as much as you can remember about what you were thinking while watching the lecture. Please feel free to pause the recording at any time to share your thoughts with me, and I will also pause it at certain points to prompt you to describe your thoughts. If you cannot recall your thought processes while watching a specific part, no worries, just say "I can't remember". Does this make sense to you? Do you have any questions before we get started?

让我们一起来看一下你在讲座之前看短片时的眼睛移动录像吧！屏幕上的这个粉红色小点显示了你正在看哪里，我们也可以听到说话的人在讲什么。但是我更想知道的是当时你脑海里在想什么。因此，我们现在将一起观看你在听讲座时的眼球移动录像。 我希望你能尽可能多地分享你在观看讲座时的想法。你可以随时暂停录制与我分享想法，我也会在某些时刻暂停录像并提示你描述当时的想法。如果你不记得在观看某个部分时的思维过程，不用担心，只需说“我不记得了”就行。你可以理解这个过程吗？在我们开始之前还有任何问题吗？

2. Instruction for researcher collecting recall data:

- What were you thinking while watching this part? 你当时听/看这个部分的时候在想什么？
- Did you encounter any difficulties while watching this part? 你在听这个部分的时

候有没有遭遇什么困难？

3. Stimulated recall data excerpts

– Participant 1 who had performed the task once

1. Researcher: What were you thinking while watching this part? 你当时听这个部分的时候在想什么?

   Participant 1: Here, I was writing down many things, like this. I wrote “cell body, soma, DNA, manufacturing”. Basically, I was writing while listening. 这里就是在写很多东西, 写 cell body, soma, DNA, manufacturing 这些。基本上都是边听边在写。

2. Researcher: What were you thinking while watching this part? 听这一段的时候，你当时在想什么？

   Participant 1: I can’t remember. 没有印象了。

3. Researcher: What were you thinking while watching this part? 听这一段的时候，你当时在想什么？

   Participant 1: At that time, I was writing this “dendrites, arbor”, and then “keep branching”. Also, I suddenly wondered if I had learned this thing in junior or high school. I was recalling its Chinese name, but I could not remember. 当时一个是记这个 dendrites, arbor, 然后 keep branching, 然后还有好像我还在想是不是在，脑子里突然在想是不是以前初高中的时候学过这个东西，在想能不能想到中文名字，但是没有想清楚。

4. Researcher: What were you thinking while listening to this part? 听这一段的时候，你当时在想什么？

   Participant 1: I was writing down “gathering information”. 就是在记 gathering information.

5. Researcher: What were you thinking while watching this part? 听这一段的时候，当时在想什么？

   Participant 1: Probably, I was taking notes. Yes, I was thinking about what she said previously. That is, “dendrite” is “local”, and this “axon” goes to different parts of the body. 应该也只是在记，对，还想了一下就是一个是她前面说的当时在想前面说的 dendrites 是 local，然后这个 axon 是到身体各个部位的。

6. Researcher: What were you thinking while listening to this part? 听到这一段的时候，当时在想什么？

   Participant 1: Here, I was thinking if this part was “synapse”. In addition, I was thinking about whether synapse referred to the gap in the middle or referred to the whole process. 这里一个就是在想是不是这个地方就是 synapse，另外一个就是在想 synapse 到底指的是中间的空隙还是整一个过程？

7. Researcher: Can you remember what you were thinking while listening to this part? 听到这里的这一段的时候有没有什么印象你当时的一个想法？

   Participant 1: No thoughts for this part. 好像这里没有。

8. Researchers: While listening to the last part, is there anything else you'd like to share about your thoughts? 听到最后这样一段的时候，对于你当时的想法还有没有什么补充的？

   Participant 1: No. To be honest, I was not concentrating while listening to the part about “muscle” and I did not take any notes. 没有，这段说实话听到那个 muscle 那边有点放空。也没有记什么笔记。

   Researchers: Why? 为什么呢？

   Participant 1: I guess I did not follow her for a moment. I heard her say “muscle”, but I was not processing what she was talking about. 应该就是有一刹那有点没跟

上，听到她说 muscle 但是脑子里没有在处理她在说什么。

– Participant 9 who had performed the task twice

1. Researcher: What were you thinking while listening to this part? 听到这一段的时候，当时在想什么？

   Participant 9: For this part, at the beginning, I first glanced at the word "dendrites" because I was not sure if I wrote it correctly. Next, I was double checking this word, and then followed what she said when she introduced this "cell body". Since I had understood this part of the information, I just listened to it. But I knew that she was going to introduce some functions of the "cell body", which I did not fully note down. Yes, that is it. 这一节就是最开始的时候，我是先瞄了一下那个 dendrites 那个词，因为我第一遍听的时候我记它记得就是, 我不确定我自己到底有没有记对。然后我就先确认一下这个词，然后接下去就是她在讲那个 cell body 的时候，就主要是跟着，因为这部分的信息，就是最开始介绍的这个信息… 这东西是什么我听得比较明确，所以就只是听一下。但是我知道后面马上就会出现关于 cell body 的一些功能，我当时没有记全。对，大概就是这样。

2. Researcher: What were you thinking while listening to this part? Can you remember ? 听到这一段的时候，你当时在想什么？还能回忆起来吗？

   Participant 9: Yes, here I was focusing on taking notes for its (soma) function. I added information about its "manufacture", and it could "produce power", and could "keep cells going", or something like this. At the beginning (of the first viewing), I took notes about the fact that it could mix all the "materials". Yes, but when I read it here, I knew it was a "cell center", but I had some doubts about this point, "makes all the materials", so I just added all the following functions to my notes. Right, so it became clearer. 对，这个地方我就是重点在… 就是在做关于它

的那个功能的 take notes，然后就是补充了一下它，关于它 manufacture，然后它可以 produce power，然后可以 keep cells going 之类的功能，因为我前面最开始的时候记过一个它是可以混合所有的 materials. 对，但是这里我自己看下来的时候，我知道它是一个 cell center，但是关于 这个 makes all the materials 就这一点我自己有点疑问，所以我就把后面的功能都补充上了，就这样。然后这样就比较清楚一点。

3. Researcher: What were you thinking while listening to this part? 听到这一段的时候，当时在想什么？

   Participant 9: The information about this part was clear to me from the first viewing. When it came to that “arbor”, I checked that piece of information again because its pronunciation was not clear to me. I then heard her say that… I could use the expression “dendritic, dendritic tree”, so I chose… I could ignore that “arbor”, and I could directly introduce the “dendrite trees”. So, I replaced “arbor” (with dendritic trees) here. 然后这个部分关于它的信息，我是第一遍就已经捕捉的比较清楚了。然后关于它那个 arbor 的时候，那个地方我又重新就是再核实了一下，因为这个发音我非常的不明确，然后我听到她说就是可以直接说 dendritic, dendritic tree, 然后我就选择… 就之前那个 arbor 我就可以忽略掉它了，我就可以直接在后面介绍里面 dendritic trees, 所以我在这个 arbor 这地方我把它就重新又替换一下 dendritic.

4. Researcher: What were you thinking while viewing this part? 听到这一段的时候，当时在想什么？

   Participant 9: During the second viewing, I was mainly focusing on “take notes”, that is, adding notes about “dendrites”, about their relationship with “information”. Because I had already written down that it branched out from the “cell body”, and had different branches. But during the second viewing, I probably focus more on its function of capturing information from externals. 第二遍主要就是在 take notes, 就是再补充一下 dendrites, 就是关于它和 information 的关系，因为我之前记的比较多的是它是从那个 cell body 延伸出来，然后会有不同的分支。但是在第二遍的时候可能会更着重于它这个从外界捕捉信息的这个功能。

5. Researcher: What about this part？那这一段呢？

   Participant 9: About this part… because she was just talking about “go for a distance”, and then I naturally looked at the far end (of the axon) following her action. Then, I additionally wrote down the information that it could “go a meter”. I “take notes” for it as well. 这个地方就是，因为她正好讲到那个 go for a distance，然后我差不多是下意识的跟着她的这个动作就延伸到了外面，然后顺便就记了一下后面它有一个那个… 它可以 go a meter 这个信息，然后就把它也 take notes 下来。

6. What were you thinking while viewing this part? 听到这一段的时候，当时在想什么？

   Participant 9: In this section, because the other information was similar to the previous information… She was talking about this “axon”. It could “carry information”. So, in this section, I was preparing for the adjective used to describe that “terminal” because it was quite unclear to me at that time. I was waiting for her to explain it. 这节的时候因为其他的信息跟前面的信息差不多，只是讲一下这个 axon 它可以就是 carry information，所以这一节我是在为后面那个 terminal 的形容词做准备，因为我当时非常的不明确，我就在等她的这个词。

   Researcher: Are you referring to this adjective? 你说的是这个形容词吗？

   Participant: Yes, it was “synaptic”. This was the word I was waiting for her to explain again. 对，就是 synaptic, 就是我在等这个词，等她再重复的解释它。

7. What were you thinking while viewing this part? 听到这一段的时候，当时在想什么？

   Participant 9: For this part, I was mainly focusing on listening to this “synapse”. That is, for this whole part, I was focusing on its spelling, pronunciation, and meaning. 这一段我就在着重听这个 synapse. 就是 整个一段就是在，关注在这个词的拼写，发音和它的意思上面。

8. What were you thinking while viewing this part? 听到这一段的时候，当时在想什么？

   Participant 9: Here, at the beginning, I glanced at the word after “cells”, but I found that it was an “or” relationship, so I thought it was okay that I did not know the meaning of the following word (dendrite). I just skipped it. And the remaining information was about the part which could interact with other cells after it arrived at the terminal. Given that it was quite clear to me during the first viewing, I just roughly verified the information during the second viewing. 在这里最开始的时候我是瞄了下 cells 后面的那个词，但是我发现它是 or 的关系，所以我觉得我后面那个词不知道也可以，我就又把它忽略掉了。然后剩下的信息就是关于它到这个终端之后，它可以和其他细胞进行交互的这个部分，因为第一遍比较明确了，所以第二遍就大概核实一下。

9. What about the final part? 那最后这一段呢？

   Participant 9: For this part, I know it was the end. There was no more meaningful information. Then, I was checking the order of my notes from the first viewing because I was thinking about how many “part” it had. During the first viewing, I wrote down these “part”. I was carefully reviewing each part and considering the sequence for retelling the lecture. 然后这个部分是因为我知道它已经结束了，没有更多的有效信息了，然后我在核实我第一遍记的笔记的那个顺序，因为我在思考后

面复述的时候… 就是它有几个 part. 然后第一遍就记了它的就是这几个 part，然后就在… 差不多就是点对点的核实一下我之后复述的时候应该用什么样的顺序比较好。

Participant 10 who had performed the task three times

1. What were you thinking while viewing this part? 听到这一段的时候，当时在想什么？

   Participant 10: Here, I was looking at where she was pointing at. She pointed at this ‘cell body’, so I looked at the ‘cell body’. Also, my eyes were scanning back and forth like this to check whether what I had written down matches these things. Then, I was checking whether the structure that she introduced matched what I wrote. 这个的话就是她指哪我看哪。她指这个 cell body, 我就看 cell body. 然后我会就是眼睛是这样来回扫，来回扫就是看一下之前写的这些东西有没有对上，然后就理一下大概她讲的这个框架跟我记得差不多，是不是差不多的。

2. (Interrupted by the participant) Then my eyes scanned the green line of that axon in the order she said. 然后我的，我的眼睛，它就这样一个顺序，我的眼睛就跟着她讲的那个顺序从就那个什么 axon 就这样顺着一条绿线就浏览了一遍。

3. What were you thinking while viewing this part? 听到这一段的时候，当时在想什么？

   Participant 10: For this part, because she said it would branch many times. So, I was looking at these branches that she was pointing at. Therefore, my eyes were moving back and forth like this, or looking down to see whether what I wrote was appropriate. I wrote down “branch”. It was a verb, so I… She said “branch” was a verb, so I put that it was a verb. I added this piece of information during the third listening because I wanted to emphasize that it was its branch, otherwise when I read them later… 这段时候是因为她说会分裂，然后分裂成各种的分支，我就看她指的这些分支，所以我的眼睛就这样动来动去的，然后或者低头看一下我写的这些东西是不是合适。我写了一个 branch，这是一个动词，所以我就… 她说的是 branch 动词，所以我就记了一个动词。然后这个是我第三遍加的，因为就想更强调一下它是它的分支，不然就可能最后再看的时候…

Researcher: What are you referring to here? 强调谁是谁的分支？

Participant 10: I was just emphasizing this “arbor” was branches of the “dendrites”. I was writing… drawing more branches so that they look more like branches. 就强调这些什么 arbor 啊什么然后 axon 是它是 这个 dendrites 的分支，想写…画更多一点就会更像它的分支。

4. While listening to this part for the third time, what were you thinking? 在听这一段的时候，你当时的想法是？第三遍时的想法。

   Participant 10: For the third time, she drew two arrows, so I was looking at the arrows. I was looking at what she was drawing. 第三遍的想法就是她画了两个箭头，我就跟着她的箭头，她画哪个我看哪个。

5. While listening to this part, what were you thinking? 在听这里这一段的时候你在想什么？

   Participant 10: While listening to this part for the third time, my eyes did not seem to be on the screen. I was just looking at this thing, thinking if it was appropriate. I was organizing what she introduced. 在听这遍的时候… 第三遍我没有，我的眼睛好像不在屏幕上。我就在看我写的这个东西是不是合适，我在理这个她讲的思路。

6. Researcher: How about this part? What were you thinking then? 在听到这一段的时候呢？你当时在想什么？

   Participant 10: While listening to this part, she was talking about the "communication center". Then, I checked my spelling. 在听这一段的时候，我在，她刚讲到的 communication center，然后，然后检查了一下我的拼写。

Researcher: Spelling of which word? 哪一个词的拼写？

Participant 10: This "synaptic". Yes, and this "synapse". 就是这个 synaptic，对，还有这个 synapse.

7. While listening to this part, what were you thinking? 在听这里这一段的时候你在想什么？

   Participant 10: While listening to this part, I wrote down "point of transformation" or something. Yes, the part that I did not write down previously. I added it to my notes. 在听这一段的时候，我记下了 point of transformation 还是什么东西。对，就是前两遍没有记下来的地方，补充一下。

8. Research: What were you thinking while listening to this final part? Can you remember? 在听这一段的时候，最后这一段的时候，你当时在想什么？还没有印象？

   Participant 10: For this final part, she… I had already written down all the key information, and then she talked about some other "cell" or "muscle". I just looked at her mouth and stared at her face. I was listening without doing anything else. 最后这一段就是她…我重点信息就都已经记到了，然后她就讲了一些什么其他的 cell 或者 muscle 之类的，就看着她的嘴，盯着她的脸，就耳朵在听，就没有其他的动作。

Appendix 12 Mixed-Effects Models for Eye-Movement Measures Extracted from the Instructor and the Diagram AOIs across Three Repetition (Research Question 1)

The instructor AOI

1. Total Fixation duration

Fixed effect	*b*	95% CI	*SE*	*df*	*t*	*p*
Intercept	3.63	[1.74, 5.51]	0.96	447.26	3.77	< .001
Time 2	0.05	[-0.01, 0.10]	0.03	17.40	1.55	0.14
Time 3	0.07	[0.01, 0.14]	0.03	25.31	2.18	0.04
Size[a]	-0.20	[-0.34, -0.06]	0.07	449.21	-2.81	.005
Duration[b]	0.69	[0.65, 0.72]	0.02	363.16	38.66	< .001
Random effect		Variance	*SD*			
Subject	(Intercept)	0.02	0.14			
	Time 2	0.14	0.38			
	Time 3	0.33	0.57			
Item	(Intercept)	0.05	0.22			
Best-fit model: FD[c] ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)						

Note. [a]Log-transformed Area of Interest size. [b]Log-transformed Area of Interest duration. [c]Log-transformed total fixation duration.

2. Fixation count

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.29	[0.26, 0.33]	0.02	14.79	< .001
Time 2	0.01	[-0.04, 0.06]	0.03	0.35	.72
Time 3	0.04	[-0.01, 0.09]	0.02	1.70	.09
Size[a]	-0.12	[-0.23, -0.02]	0.05	-2.36	.02
Duration[b]	0.30	[0.28, 0.33]	0.01	24.71	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.004	0.06		
Item	(Intercept)	0.002	0.05		
Best-fit model: FC[b] ~ Time + Size + Duration + (1 | Subject) + (1 | Instance)					

Note. [a]Log-transformed Area of Interest size. [b]Log-transformed Area of Interest duration. [c]Fixation count.

3. Skip rate

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	2.01	[1.77, 2.26]	0.13	15.83	< .001
Time2	0.29	[0.04, 0.54]	0.13	2.28	.02
Time3	0.04	[-0.42, 0.51]	0.24	0.18	.86
Size[a]	0.42	[0.03, 0.80]	0.20	2.13	.03
Duration[b]	-0.81	[-0.91, -0.70]	0.05	-15.57	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.42	0.65		
	Time2	0.45	0.67		
	Time3	1.63	1.28		
Item	(Intercept)	0.69	0.83		
Best-fit model: SR[c] ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)					

Note. [a]Log-transformed Vocabulary Size Test scores. [b]Log-transformed Area of Interest size. [c]AOI instance skip rate.

The Diagram AOI

1. Total Fixation duration

Fixed effect	*b*	95% CI	*SE*	*df*	*t*	*p*
Intercept	0.77	[1.74, 5.51]	0.35	520.12	2.18	.03
Time 2	-0.06	[-0.01, 0.10]	0.03	26.58	-2.13	.04
Time 3	-0.05	[0.01, 0.14]	0.02	22.56	-2.31	.03
Size[a]	0.05	[-0.34, -0.06]	0.03	512.34	1.78	.08
Duration[b]	0.65	[0.65, 0.72]	0.01	419.16	63.94	< .001
Random effect		Variance	*SD*			
Subject	(Intercept)	0.02	0.14			
	Time 2	0.14	0.38			
	Time 3	0.33	0.57			
Item	(Intercept)	0.05	0.22			
Best-fit model: FD[c] ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)						

Note. [a]Log-transformed Area of Interest size. [b]Log-transformed Area of Interest duration. [c]Log-transformed total fixation duration.

2. Fixation count

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.40	[0.38, 0.43]	0.01	29.93	< .001
Time 2	-0.002	[-0.03, 0.03]	0.02	-0.12	.91
Time 3	-0.04	[-0.07, -0.01]	0.02	-2.32	.02
Size[a]	0.08	[0.03, 0.12]	0.02	3.54	< .001
Duration[b]	0.33	[0.31, 0.35]	0.01	40.51	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.002	0.04		
Item	(Intercept)	0.003	0.06		
Best-fit model: FC[c] ~ Time + Size + Duration + (1 | Subject) + (1 | Instance)					

Note. [a]Log-transformed Area of Interest size. [b]Log-transformed Area of Interest duration. [c]Fixation count.

3. Skip rate

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.40	[0.19, 0.60]	0.10	3.87	< .001
Time2	0.53	[0.23, 0.82]	0.15	3.52	< .001
Time3	0.52	[0.04, 0.99]	0.24	2.14	.03
Size[a]	-0.61	[-0.77, -0.45]	0.08	-7.57	< .001
Duration[b]	-0.28	[-0.35, -0.21]	0.03	-8.37	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.29	0.54		
	Time2	0.65	0.81		
	Time3	1.71	1.31		
Item	(Intercept)	0.28	0.53		
Best-fit model: SR[c] ~ Time + Size + Duration + (1 + Time | Subject) + (1 | Instance)					

Note. [a]Log-transformed Area of Interest size. [b]Log-transformed Area of Interest duration. [c]AOI instance skip rate.

4. Interactive saccades between the instructor AOI and the diagram AOI

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	4.28	[4.15, 4.40]	0.06	67.05	< .001
Time2	-0.24	[-0.30, -0.18]	0.03	-7.62	< .001
Time3	-0.01	[-0.07, 0.05]	0.03	-0.28	0.78
Random effect		Variance	*SD*		
Subject	(Intercept)	0.11	0.33		
Best-fit model: saccade[a] ~ Time + (1 | Subject)					

Note. [a]The number of integrative saccades between the instructor AOI and the diagram AOIs.

Appendix 13 Mixed-Effects Models of the Relationship between Eye-Movement Measures and Vocabulary Post-Tests Scores (Research Question 3)

Control Group

1. Percentage of correct idea units – Total fixation duration on the instructor AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	130.83	[72.10, 189.56]	28.67	4.56	< .001
FD[a]	-11.03	[-17.06, -5.00]	2.94	-3.75	< .001
Linear regression model: correctIU[b] ~ FD					

Note. [a]Log-transformed total fixation duration on the instructor AOI.
[b]The percentage of correct idea units.

2. Percentage of correct idea units – Total fixation duration on the diagram AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	171.41	[-7.44, 350.26]	87.31	1.96	.06
FD[a]	-13.64	[-30.13, 2.85]	8.05	-1.69	.10
Linear regression model: correctIU[b] ~ FD					

Note. [a]Log-transformed total fixation duration on the diagram AOI.
[b]The percentage of correct idea units.

3. Percentage of incorrect idea units – Total fixation duration on the instructor AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	6.92	[-10.74, 24.59]	8.62	0.80	.43
FD[a]	-0.29	[-2.10, 1.53]	0.89	-0.32	.75
Linear regression model: incorrectIU[b] ~ FD					

Note. [a]Log-transformed total fixation duration on the instructor AOI.
[b]The percentage of incorrect idea units.

4. Percentage of incorrect idea units – Total fixation duration on the diagram AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	14.65	[-31.35, 60.64]	22.45	0.65	.52
FD[a]	-0.97	[-5.21, 3.27]	2.07	-0.47	.64
Linear regression model: incorrectIU[b] ~ FD					

Note. [a]Log-transformed total fixation duration on the instructor AOI.
[b]The percentage of incorrect idea units.

Repetition Group

1. Percentage of correct idea units – Total fixation duration on the instructor AOI at Time 1

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	24.50	[-54.59, 103.59]	38.61	0.63	.53
FD1[a]	1.54	[-6.36, 9.54]	3.91	0.39	.70
Linear regression model: correctIU[b] ~ FD1					

Note. [a]Log-transformed total fixation duration on the instructor AOI at Time 1.
[b]The percentage of correct idea units.

2. Percentage of correct idea units – Total fixation duration on the instructor AOI at Time 2

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	97.88	[44.12, 151.64]	26.24	3.73	< .001
FD2[a]	-6.05	[-11.61, -0.48]	2.72	-2.23	.03
Linear regression model: correctIU[b] ~ FD2					

Note. [a]Log-transformed total fixation duration on the instructor AOI at Time 2.
[b]The percentage of correct idea units.

3. Percentage of correct idea units – Total fixation duration on the instructor AOI at Time 3

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	68.76	[47.66, 89.87]	10.30	6.68	< .001
FD3[a]	-3.01	[-5.14, -0.87]	1.04	-2.89	.007
Linear regression model: correctIU[b] ~ FD3					

Note. [a]Log-transformed total fixation duration on the instructor AOI at Time 3.
[b]The percentage of correct idea units.

4. Percentage of correct idea units – Cumulative fixation duration on the instructor AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	119.42	[34.80, 204.04]	41.31	2.89	.007
CFD[a]	-7.24	[-14.91, 0.43]	3.75	-1.93	.06
Linear regression model: correctIU[b] ~ CFD					

Note. [a]Log-transformed cumulative fixation duration on the instructor AOI.
[b]The percentage of correct idea units.

5. Percentage of correct idea units – Total fixation duration on the diagram AOI at time 1

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	51.00	[-105. 38, 207.39]	76.35	0.67	.51
FD1[a]	-1.05	[-15.58, 13.48]	7.09	-0.1	.88
Linear regression model: correctIU[b] ~ FD1					

Note. [a]Log-transformed total fixation duration on the diagram AOI at Time 1.
[b]The percentage of correct idea units.

6. Percentage of correct idea units – Total fixation duration on the diagram AOI at time 2

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	117.11	[18.35, 215.86]	48.21	2.43	.02
FD2[a]	-7.44	[-16.93, 2.04]	4.63	-1.61	.12
Linear regression model: correctIU[b] ~ FD2					

Note. [a]Log-transformed total fixation duration on the diagram AOI at Time 2.
[b]The percentage of correct idea units.

7. Percentage of correct idea units – Total fixation duration on the diagram AOI at time 3

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	86.18	[42.43, 129.93]	21.36	4.04	< .001
FD3[a]	-4.54	[-8,79, -0.29]	2.07	-2.19	.04
Linear regression model: correctIU[b] ~ FD3					

Note. [a]Log-transformed total fixation duration on the diagram AOI at Time 3.
[b]The percentage of correct idea units.

8. Percentage of correct idea units – Cumulative fixation duration on the diagram AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	171.84	[3.41, 340. 27]	82.23	2.09	.05
CFD[a]	-11.32	[-25.74, 3.10]	7.04	-1.61	.12
Linear regression model: correctIU[b] ~ CFD					

Note. [a]Log-transformed cumulative fixation duration on the diagram AOI.
[b]The percentage of correct idea units.

9. Percentage of incorrect idea units – Total fixation duration on the instructor AOI at time 1

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	5.73	[-9,71, 21. 19]	7.54	0.76	.45
FD1[a]	-0.33	[-1.90, 1.23]	0.76	-0.44	.67
Linear regression model: incorrectIU[b] ~ FD1					

Note. [a]Log-transformed total fixation duration on the instructor AOI at Time 1.
[b]The percentage of incorrect idea units.

10. Percentage of incorrect idea units – Total fixation duration on the instructor AOI at time 2

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	4.89	[-6.47, 16.25]	5.54	0.88	.39
FD2[a]	-0.25	[-1.43, 0.92]	0.57	-0.44	.66
Linear regression model: incorrectIU[b] ~ FD2					

Note. [a]Log-transformed total fixation duration on the instructor AOI at Time 2.
[b]The percentage of incorrect idea units.

11. Percentage of incorrect idea units – Total fixation duration on the instructor AOI at time 3

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	1.83	[-2.86, 6.53]	2.29	0.80	.43
FD3[a]	0.06	[-0.41, 0.54]	0.23	0.27	.79
Linear regression model: incorrectIU[b] ~ FD3					

Note. [a]Log-transformed total fixation duration on the instructor AOI at Time 3.
[b]The percentage of incorrect idea units.

12. Percentage of incorrect idea units – Cumulative fixation duration on the instructor AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	4.84	[-12.75, 22.42]	8.58	0.56	.58
CFD[a]	-0.22	[-1.81, 1.38]	0.79	-0.28	.78
Linear regression model: incorrectIU[b] ~ CFD					

Note. [a]Log-transformed cumulative fixation duration on the instructor AOI.
[b]The percentage of incorrect idea units.

13. Percentage of incorrect idea units – Total fixation duration on the diagram AOI at time 1

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	4.74	[-25.83, 35.30]	14.92	0.32	.75
FD1[a]	-0.21	[-3.05, 2.63]	1.39	-0.15	.88
Linear regression model: incorrectIU[b] ~ FD1					

Note. [a]Log-transformed total fixation duration on the diagram AOI at Time 1.
[b]The percentage of incorrect idea units.

14. Percentage of incorrect idea units – Total fixation duration on the diagram AOI at time 2

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	-5.19	[-25.15, 14.76]	9.74	-0.53	0.60
FD2[a]	0.73	[-1.18, 2.65]	0.94	0.79	0.44
Linear regression model: incorrectIU[b] ~ FD2					

Note. [a]Log-transformed total fixation duration on the diagram AOI at Time 2.
[b]The percentage of incorrect idea units.

15. Percentage of incorrect idea units – Total fixation duration on the diagram AOI at time 3

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	1.07	[-8.17, 10.31]	4.51	0.24	.81
FD3[a]	0.13	[-0.76, 1.03]	0.44	0.31	.76
Linear regression model: incorrectIU[b] ~ FD3					

Note. [a]Log-transformed total fixation duration on the diagram AOI at Time 3.
[b]The percentage of incorrect idea units.

16. Percentage of incorrect idea units – Cumulative fixation duration on the diagram AOI

Fixed effect	*b*	95% CI	*SE*	*t*	*p*
Intercept	-5.19	[-39.47, 29.09]	16.73	-0.31	.76
CFD[a]	0.65	[-2.28, 3.59]	1.43	0.46	.65
Linear regression model: incorrectIU[b] ~ CFD					

Note. [a]Log-transformed cumulative fixation duration on the diagram AOI.
[b]The percentage of incorrect idea units.

Appendix 14 Mixed-Effects Models of Learning Gains in the Immediate and Delayed Vocabulary Post-Tests (Research Question 4)

1. Immediate form recognition

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	1.33	[0.79, 1.87]	0.27	4.90	< .001
Group2[a]	0.53	[0.07, 1.00]	0.24	2.24	.03
Random effect		Variance	*SD*		
Subject	(Intercept)	0.25	0.50		
Item	(Intercept)	1.18	1.09		
	Group2	0.29	0.54		
Best-fit model: FReco1[b] ~ Group + (1 | Subject) + (1 + Group | Item)					

Note. [a]Repetition group. [b]Immediate form recognition test scores.

2. Immediate meaning recognition

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-1.29	[-2.44, -0.14]	0.59	-2.19	.03
Group2[a]	1.36	[0.83, 1.90]	0.27	5.00	< .001
CAE[b]	0.07	[0.01, 0.13]	0.03	2.45	.01
POS[c]	1.19	[0.48, 1.90]	0.36	3.28	.001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.49	0.70		
Item	(Intercept)	0.17	0.41		
Best-fit model: MReco1[d] ~ Group + CAE + POS + (1 | Subject) + (1 | Item)					

Note. [a]Repetition group. [b]Cambridge Certificate in Advanced English test scores. [c]Part of speech. [d]Immediate meaning recognition test scores.

3. Immediate meaning recall

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-1.96	[-2.48, -1.45]	0.26	-7.48	< .001
Group2[a]	1.45	[0.88, 2.03]	0.29	4.94	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.86	0.93		
Item	(Intercept)	0.33	0.57		
	Group2	0.15	0.39		
Best-fit model: MReca1[b] ~ Group + (1 | Subject) + (1 + Group| Item)					

Note. [a]Repetition group. [b]Immediate meaning recall test scores.

4. Delayed form recognition

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	1.52	[0.92, 2.12]	0.31	4.97	< .001
Group2[a]	0.71	[0.29, 1.12]	0.21	3.34	.001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.29	0.54		
Item	(Intercept)	1.54	1.24		
Best-fit model: FReco2[b] ~ Group + (1 + | Subject) + (1 + | Item)					

Note. [a]Repetition group. [b]Delayed form recognition test scores.

5. Delayed meaning recognition

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-1.46	[-2.62, -0.31]	0.59	-2.48	.01
Group2[a]	0.86	[0.34, 1.39]	0.27	3.21	.001
CAE[b]	0.08	[0.02, 0.13]	0.03	2.67	.01
POS[c]	1.20	[0.48, 1.92]	0.37	3.26	.001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.56	0.75		
Item	(Intercept)	0.19	0.43		
Best-fit model: MReco2[d] ~ Group + CAE + POS + (1 | Subject) + (1 | Item)					

Note. [a]Repetition group. [b]Cambridge Certificate in Advanced English test scores.
[c]Part of speech. [d]Delayed Meaning recognition test scores.

6. Delayed meaning recall

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-5.00	[-6.27, -3.74]	0.64	-7.78	< .001
Group2[a]	1.46	[0.75, 2.18]	0.36	4.01	< .001
CAE[b]	0.11	[0.05, 0.17]	0.03	3.73	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	0.88	0.94		
Item	(Intercept)	0.39	0.63		
	Group2	0.50	0.71		
Best-fit model: MReca2[c] ~ Group + CAE + (1 | Subject) + (1 + Group| Item)					

Note. [a]Repetition group. [b]Cambridge Certificate in Advanced English test scores. [c]Delayed meaning recall test scores.

Appendix 15 Mixed-Effects Models for Eye-Movement Measures Extracted from the Target Word AOIs across Three Repetition (Research Question 6)

1. Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*df*	*t*	*p*
Intercept	-5.58	[-11.31, -0.15]	2.92	8.03	-1.91	.09
Time 2	-0.49	[-0.66, -0.32]	0.09	28.46	-5.51	< .001
Time 3	-0.59	[-0.83, -0.36]	0.12	27.45	-4.98	< .001
Size[a]	1.39	[0.78, 1.99]	0.31	8.02	4.52	.002
Nonverbal[b]	0.11	[0.03, 0.19]	0.04	7.97	2.86	.02
Random effect		Variance	*SD*			
Subject	(Intercept)	0.08	0.28			
	Time 2	0.14	0.38			
	Time 3	0.33	0.57			
Item	(Intercept)	0.12	0.35			
Best-fit model: FD[c] ~ Time + Size + Nonverbal + (1 + Time | Subject) + (1 | Item)						

Note. [a]Log-transformed Area of Interest size. [b]Number of nonverbal signals.
[c]Log-transformed total fixation duration.

2. Mean fixation duration

Fixed effect	*b*	95% CI	*SE*	*df*	*t*	*p*
Intercept	5.49	[5.32, 5.67]	0.09	10.93	61.90	< .001
Time 2	-0.07	[-0.12, -0.02]	0.03	882.77	-2.56	.01
Time 3	-0.09	[-0.14, -0.03]	0.03	885.17	-3.12	.002
Length[a]	0.03	[0.00, 0.05]	0.01	9.09	2.26	.05
Random effect		Variance	*SD*			
Subject	(Intercept)	0.01	0.12			
Item	(Intercept)	0.01	0.08			
Best-fit model: MFD[b] ~ Time + Length + (1 | Subject) + (1 | Item)						

Note. [a]Length of target items. [b]Log-transformed mean fixation duration.

3. Fixation count

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-11.80	[-17.93, -5.66]	3.13	-3.77	< .001
Time 2	-0.38	[-0.50, -0.26]	0.06	-6.24	< .001
Time 3	-0.47	[-0.65, -0.30]	0.09	-5.23	< .001
Size[a]	1.48	[0.84, 2.11]	0.32	4.56	< .001
Random effect		Variance	*SD*		
Obs_effect	(Intercept)	0.13	0.36		
Subject	(Intercept)	0.05	0.21		
	Time 2	0.07	0.26		
	Time 3	0.19	0.44		
Item	(Intercept)	0.16	0.40		
Best-fit model: FC[b] ~ Time + Size+ (1 | obs_effect) + (1 + Time | Subject) + (1 | Item)					

Note. [a]Log-transformed Area of Interest size. [b]Fixation count.

4. Run count

Fixed effect	*b*	*95% CI*	*SE*	*z*	*p*
Intercept	-10.11	[-16.16, -4.07]	3.09	-3.28	.001
Time 2	-0.29	[-0.40, -0.18]	0.06	-5.08	<.001
Time 3	-0.38	[-0.53, -0.22]	0.08	-4.81	<.001
Size[a]	1.27	[0.64, 1.89]	0.32	3.96	<.001
Random effect		Variance	SD		
Obs_effect	(Intercept)	0.06	0.25		
Subject	(Intercept)	0.04	0.21		
	Time 2	0.06	0.24		
	Time 3	0.14	0.38		
Item	(Intercept)	0.16	0.40		
Best-fit model: RC[b] ~ Time + Size + (1 | obs_effect) + (1 + Time | Subject) + (1 | Item)					

Note. [a]Log-transformed Area of Interest size. [b]Run count.

5. Skip rate

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-27.41	[-58.94, 4.12]	16.09	-1.70	.09
Time2	1.49	[0.33, 2.64]	0.59	2.52	.01
Time3	3.00	[1.90, 4.10]	0.56	5.35	< .001
VST[a]	4.74	[1.52, 7.96]	1.64	2.89	.004
Size[b]	-2.12	[-3.54, -0.70]	0.72	-2.93	.003
Nonverbal[c]	-0.40	[-0.73, -0.06]	0.17	-2.31	.02
Random effect		Variance	*SD*		
Subject	(Intercept)	0.61	0.78		
Item	(Intercept)	0.29	0.54		
Best-fit model: SR[d]~ Time + VST + Size + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed Vocabulary Size Test scores. [b]Log-transformed Area of Interest size. [c]Number of nonverbal signals. [d]AOI skip rate.

Appendix 16 Mixed-Effects Models of the Relationship between Eye-Movement Measures and Vocabulary Post-Tests Scores (Research Question 7)

Control group

1. Immediate form recognition – Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.13	[-2.24, 2.49]	1.21	0.11	.91
FD[a]	0.19	[-0.10, 0.48]	0.15	1.26	.21
Random effect		Variance	*SD*		
Subject	(Intercept)	0.10	0.31		
Item	(Intercept)	0.37	0.61		
Best-fit model: FReco1[b] ~ FD + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration. [b]Immediate form recognition test scores.

2. Immediate meaning recognition – Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.63	[-2.53, 1.32]	0.96	-0.66	.51
FD[a]	0.09	[-0.16, 0.32]	0.12	0.71	.48
POS[b]	1.05	[0.38, 1.79]	0.34	3.14	.002
Random effect		Variance	*SD*		
Subject	(Intercept)	0.37	0.61		
Item	(Intercept)	0.04	0.20		
Best-fit model: MReco1[c] ~ FD + POS + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration. [b]Part of speech.
[c]Immediate meaning recognition test scores.

3. Immediate meaning recall – Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-2.74	[-4.73, -0.84]	0.96	-2.85	.004
FD[a]	0.10	[-0.13, 0.33]	0.12	0.83	.40
Random effect		Variance	*SD*		
Subject	(Intercept)	0.82	0.91		
Item	(Intercept)	0.32	0.57		
Best-fit model: MReca1[c] ~ FD + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration. [b]Immediate meaning recall test scores.

4. Delayed form recognition – Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	4.73	[0.50, 8.96]	2.16	2.19	.03
FD[a]	-0.29	[-0.83, 0.24]	0.27	-1.07	.28
Nonverbal[b]	0.71	[0.18, 1.23]	0.27	2.65	.01
Random effect		Variance	*SD*		
Subject	(Intercept)	3.37	1.84		
Item	(Intercept)	0.31	0.55		
Best-fit model: FReco2[c] ~ FD + Nonverbal + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration. [b]Number of nonverbal signals.
[c]Delayed form recognition test scores.

5. Delayed meaning recognition – Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-1.03	[-3.11, 1.08]	1.05	-0.98	.33
FD[a]	0.13	[-0.13, 0.39]	0.13	1.01	.31
POS[b]	0.88	[0.08, 1.75]	0.39	2.28	.02
Random effect		Variance	*SD*		
Subject	(Intercept)	0.42	0.65		
Item	(Intercept)	0.12	0.35		
Best-fit model: MReco2[c] ~ FD + POS + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration. [b]Part of speech.
[c]Delayed meaning recognition test scores.

6. Delayed meaning recall – Total fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-7.50	[-10.97, -4.39]	1.65	-4.55	< .001
FD[a]	0.19	[-0.11, 0.53]	0.16	1.20	.23
CAE[b]	0.14	[0.03, 0.24]	0.05	2.67	.01
POS[c]	1.06	[0.22, 1.97]	0.40	2.67	.01
Random effect		Variance	*SD*		
Subject	(Intercept)	1.09	1.04		
Item	(Intercept)	0.17	0.42		
Best-fit model: MReca2[d] ~ FD + CAE + POS + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration.
[b]Cambridge Certificate in Advanced English test scores.
[c]Part of speech. [d]Delayed meaning recall test scores.

Repetition Group

1. Immediate form recognition – Total fixation duration at Time 1

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.12	[-2.73, 2.49]	1.33	-0.09	.93
FD1[a]	0.25	[-0.07, 0.57]	0.17	1.52	.13
Random effect		Variance	*SD*		
Subject	(Intercept)	0.10	0.31		
Item	(Intercept)	0.12	0.34		
Best-fit model: FReco1[b] ~ FD1 + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 1.
[b]Immediate form recognition test scores.

2. Immediate meaning recognition – Total fixation duration at Time 1

Fixed effect		*b*	95% CI	*SE*	*z*	*p*
Intercept		-0.95	[-4.15, 2.26]	1.64	-0.58	.56
FD1[a]		0.03	[-0.33, 0.40]	0.19	0.18	.86
CAE[b]		0.10	[0.01, 0.19]	2.14	2.14	.03
Nonverbal[c]		0.24	[0.05, 0.42]	2.49	2.49	.01
Random effect			Variance	*SD*		
Subject	(Intercept)		0.61	0.78		
Item	(Intercept)		0.15	0.39		
Best-fit model: MReco1[d] ~ FD1 + CAE + Nonverbal + (1 | Subject) + (1 | Item)						

Note. [a]Log-transformed total fixation duration at Time 1.
[b]Cambridge Certificate in Advanced English test scores.
[c]Number of nonverbal signals. [d]Immediate meaning recognition test scores.

3. Immediate meaning recall – Total fixation duration at Time 1

Fixed effect		*b*	95% CI	*SE*	*z*	*p*
Intercept		5.26	[-0.61, 11.13]	3.00	1.75	.08
FD1[a]		0.19	[-0.05, 0.42]	0.12	1.55	.12
Size [b]		-0.76	[,]	0.34	-2.23	.03
Random effect			Variance	*SD*		
Subject	(Intercept)		9.25	3.04		
	FD1		0.08	0.28		
Item	(Intercept)		0.09	0.30		
Best-fit model: MReca1[c] ~ FD1 + (1 + FD1 | Subject) + (1 | Item)						

Note. [a]Log-transformed total fixation duration at Time 1.
[b]Immediate meaning recall test scores.

4. Delayed form recognition – Total fixation duration at Time 1

Fixed effect		*b*	95% CI	*SE*	*z*	*p*
Intercept		1.58	[-2.13, 5.30]	1.90	0.84	.40
FD1[a]		0.21	[-0.24, 0.67]	0.23	0.91	.36
Random effect			Variance	*SD*		
Subject	(Intercept)		1.06	1.03		
Best-fit model: FReco2[b] ~ FD1 + (1 | Subject)						

Note. [a]Log-transformed total fixation duration at Time 1.
[b]Delayed form recognition test scores.

5. Delayed meaning recognition – Total fixation duration at Time 1

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.56	[-2.15, 3.27]	1.38	0.41	.69
FD1[a]	-0.02	[-0.37, 0.33]	0.18	-0.10	.92
Nonverbal[b]	0.33	[0.14, 0.51]	0.10	3.42	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	1.08	1.04		
Item	(Intercept)	0.14	0.37		
Best-fit model: MReco2[c] ~ FD1 + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 1. [b]Number of nonverbal signals. [c]Delayed meaning recognition test scores.

6. Delayed meaning recall – Total fixation duration at Time 1

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-3.35	[-5.54, -1.15]	1.12	-2.99	.003
FD1[a]	0.23	[-0.03, 0.49]	0.13	1.71	.09
Random effect		Variance	*SD*		
Subject	(Intercept)	0.77	4.74		
Item	(Intercept)	0.12	0.35		
Best-fit model: MReca2[b] ~ FD1 + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 1.
[b]Delayed meaning recall test scores.

7. Immediate form recognition – Total fixation duration at Time 2

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	1.26	[-1.06, 3.58]	1.18	1.06	.29
FD2[a]	0.09	[-0.21, 0.40]	0.15	0.59	.55
Random effect		Variance	*SD*		
Subject	(Intercept)	0.10	0.32		
Item	(Intercept)	0.07	0.26		
Best-fit model: FReco1[b] ~ FD2 + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 2.
[b]Immediate form recognition test scores.

8. Immediate meaning recognition – Total fixation duration at Time 2

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	1.60	[-0.79, 3.99]	1.22	1.32	.19
FD2[a]	-0.08	[-0.40, 0.25]	0.17	-0.45	.65
Nonverbal[b]	0.26	[0.08, 0.44]	0.09	2.81	.005
Random effect		Variance	*SD*		
Subject	(Intercept)	0.74	0.86		
Item	(Intercept)	0.06	0.25		
Best-fit model: MReco1[c] ~ FD2 + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 2.
[c]Number of nonverbal signals. [c]Immediate meaning recognition test scores.

9. Immediate meaning recall – Total fixation duration at Time 2

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-1.71	[-3.53, 0.10]	0.92	-1.86	.06
FD2[a]	-0.01	[-0.18, 0.16]	0.09	-0.13	.90
CAE[b]	0.07	[0.01, 0.13]	0.03	2.16	.03
Random effect		Variance	*SD*		
Subject	(Intercept)	0.66	0.81		
Item	(Intercept)	0.10	0.31		
Best-fit model: MReca1[c] ~ FD2 + CAE + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time2.
[b]Cambridge Certificate in Advanced English test scores.
[c]Immediate meaning recall test scores.

10. Delayed form recognition – Total fixation duration at Time 2

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	2.67	[-0.93, 6.28]	1.84	1.45	.15
FD2[a]	0.08	[-0.39, 0.55]	0.24	0.34	.74
Random effect		Variance	*SD*		
Subject	(Intercept)	0.70	0.83		
Best-fit model: FReco2[b] ~ FD2 + (1 | Subject)					

Note. [a]Log-transformed total fixation duration at Time 2.
[b]Delayed form recognition test scores.

11. Delayed meaning recognition – Total fixation duration at Time 2

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	1.92	[-0.51, 4.34]	1.24	1.55	.12
FD2[a]	-0.21	[-0.54, 0.12]	0.17	-1.25	.21
Nonverbal[b]	0.37	[0.18, 0.56]	0.10	3.75	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	1.19	1.09		
Item	(Intercept)	0.15	0.39		
Best-fit model: MReco2[c] ~ FD2 + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 2. [b]Number of nonverbal signals. [c]Delayed meaning recognition test scores.

12. Delayed meaning recall – Total fixation duration at Time 2

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.67	[-1.79, 0.45]	0.57	-1.17	.24
FD2[a]	-0.11	[-0.25, 0.03]	0.07	-1.49	.14
Random effect		Variance	*SD*		
Subject	(Intercept)	0.72	0.85		
Best-fit model: MReca2[b] ~ FD2 + (1 | Subject)					

Note. [a]Log-transformed total fixation duration at Time 2.
[b]Delayed meaning recall test scores.

13. Immediate form recognition – Total fixation duration at Time 3

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.60	[-1.82, 3.01]	1.23	0.49	.63
FD3[a]	0.17	[-0.14, 0.49]	0.16	1.08	.28
Random effect		Variance	*SD*		
Subject	(Intercept)	0.13	0.36		
Item	(Intercept)	0.17	0.42		
Best-fit model: FReco1[b] ~ FD3 + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 3.
[b]Immediate form recognition test scores.

14. Immediate meaning recognition – Total fixation duration at Time 3

Fixed effect	b	95% CI	*SE*	*z*	*p*
Intercept	2.23	[-0.30, 4.76]	1.29	1.73	.08
FD3[a]	-0.11	[-0.44, 0.22]	0.17	-0.66	.51
POS[b]	1.22	[0.26, 2.18]	0.49	2.49	.01
Random effect		Variance	*SD*		
Subject	(Intercept)	0.59	0.77		
Item	(Intercept)	0.08	0.29		
Best-fit model: MReco1[c] ~ FD3 + POS + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 3. [b]Part of speech.
[c]Immediate form recognition test scores.

15. Immediate meaning recall – Total fixation duration at Time 3

Fixed effect	b	95% CI	*SE*	*z*	*p*
Intercept	-0.25	[-1.53, 1.02]	0.64	-0.39	.70
FD3[a]	-0.04	[-0.20, 0.12]	0.08	0.45	.65
Random effect		Variance	*SD*		
Subject	(Intercept)	0.75	0.87		
Item	(Intercept)	0.09	0.29		
Best-fit model: MReca1[b] ~ FD3 + (1 | Subject) + (1 + | Item)					

Note. [a]Log-transformed total fixation duration at Time 3.
[b]Immediate meaning recall test scores.

16. Delayed form recognition – Total fixation duration at Time 3

Fixed effect	b	95% CI	*SE*	*z*	*p*
Intercept	1.13	[-2.58, 4.84]	1.89	0.60	.55
FD3[a]	0.32	[-0.17, 0.81]	0.25	1.27	.20
Random effect		Variance	*SD*		
Subject	(Intercept)	1.24	1.11		
Best-fit model: FReco2[b] ~ FD3 + (1 | Subject)					

Note. [a]Log-transformed total fixation duration at Time 3.
[b]Delayed form recognition test scores.

17. Delayed meaning recognition – Total fixation duration at Time 3

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	0.81	[-1.89, 3.52]	1.38	0.59	.56
FD3[a]	-0.07	[-0.44, 0.29]	0.19	-0.40	.69
Nonverbal[b]	0.38	[0.15, 0.61]	0.12	3.22	.001
Random effect		Variance	*SD*		
Subject	(Intercept)	1.45	1.20		
Item	(Intercept)	0.42	0.65		
Best-fit model: MReco2[c] ~ FD3 + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 3. [b]Number of nonverbal signals. [c]Delayed meaning recognition test scores.

18. Delayed meaning recall – Total fixation duration at Time 3

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.90	[-2.17, 0.37]	0.65	-1.39	.16
FD3[a]	-0.09	[-0.25, 0.08]	0.08	-1.03	.30
Random effect		Variance	*SD*		
Subject	(Intercept)	0.59	0.77		
Item	(Intercept)	0.02	0.15		
Best-fit model: MReca2[b] ~ FD3 + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed total fixation duration at Time 3.
[b]Delayed meaning recall test scores.

19. Immediate form recognition – Cumulative fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.32	[-3.72, 3.09]	1.74	-0.18	.86
CFD[a]	0.25	[-0.13, 0.64]	0.20	1.29	.20
Random effect		Variance	*SD*		
Subject	(Intercept)	0.10	0.32		
Item	(Intercept)	0.15	0.39		
Best-fit model: FReco1[b] ~ CFD + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed cumulative fixation duration.
[b]Immediate form recognition test scores.

20. Immediate meaning recognition – Cumulative fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.99	[-5.41, 3.03]	2.10	-0.47	.64
CFD[a]	0.04	[-0.39, 0.51]	0.22	0.17	.87
CAE[b]	0.23	[0.01, 0.20]	0.10	2.35	.02
Nonverbal[c]	0.10	[0.04, 0.44]	0.04	2.15	.03
Random effect		Variance	*SD*		
Subject	(Intercept)	0.61	0.78		
Item	(Intercept)	0.15	0.39		
Best-fit model: MReco1[d] ~ CFD + CAE + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed cumulative fixation duration.
[b]Cambridge Certificate in Advanced English test scores.
[c]Number of nonverbal signals. [d]Immediate meaning recognition test scores.

21. Immediate meaning recall – Cumulative fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.77	[-4.73, -0.84]	1.42	-0.55	.59
CFD[a]	0.03	[-0.13, 0.33]	0.15	0.19	.85
Random effect		Variance	*SD*		
Subject	(Intercept)	17.49	4.18		
	CFD	0.15	0.38		
Item	(Intercept)	0.13	0.36		
Best-fit model: MReca1[b] ~ CFD + (1 + CFD | Subject) + (1 | Item)					

Note. [a]Log-transformed cumulative fixation duration.
[b]Immediate meaning recall test scores.

22. Delayed form recognition – Cumulative fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-0.24	[-4.44, 3.97]	2.14	-0.11	.91
CFD[a]	0.40	[-0.08, 0.88]	0.25	1.63	.10
Random effect		Variance	*SD*		
Subject	(Intercept)	1.11	1.05		
Best-fit model: FReco2[b] ~ CFD + (1 | Subject)					

Note. [a]Log-transformed cumulative fixation duration.
[b]Delayed form recognition test scores.

23. Delayed meaning recognition – Cumulative fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	3.08	[-0.47, 6.76]	1.79	1.71	.09
CFD[a]	-0.31	[-0.74, 0.10]	0.21	-1.50	.13
Nonverbal[b]	0.38	[0.20, 0.60]	0.10	3.86	< .001
Random effect		Variance	*SD*		
Subject	(Intercept)	1.12	1.06		
Item	(Intercept)	0.12	0.35		
Best-fit model: MReco2[c] ~ CFD + Nonverbal + (1 | Subject) + (1 | Item)					

Note. [a]Log-transformed cumulative fixation duration. [b]Number of nonverbal signals. [c]Delayed meaning recognition test scores.

24. Delayed meaning recall – Cumulative fixation duration

Fixed effect	*b*	95% CI	*SE*	*z*	*p*
Intercept	-2.46	[-5.70, 0.18]	1.42	-1.74	.08
CFD[a]	0.10	[-0.17, 0.44]	0.15	0.72	.47
Random effect		Variance	*SD*		
Subject	(Intercept)	22.51	4.74		
	CFD	0.19	0.44		
Item	(Intercept)	0.02	0.15		
Best-fit model: MReca2[b] ~ CFD + (1 + CFD | Subject) + (1 | Item)					

Note. [a]Log-transformed cumulative fixation duration. [b]Delayed meaning recall test scores.