Incidental Vocabulary Acquisition from Synchronous Reading-while-Listening vs. Asynchronous Reading-and-Listening: a Randomised Trial.

Christopher Lack

MSc in Applied Linguistics for Language Teaching

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION



1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].

3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.

4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.

5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Christopher Lack
Date:	22nd September 2023

Abstract

Incidental vocabulary learning appears to be crucial for language acquisition. Vocabulary can be learnt incidentally from reading, but this is a slow process, so scholars have sought techniques to speed it up. One such technique is simultaneously reading-while-listening to a text, which appears to be a more effective way of facilitating incidental vocabulary learning. However, cognitive load theory suggests that reading-while-listening may not be an optimal approach, because it imposes a heavy load on working memory. An alternative method, asynchronous reading-and-listening, may be less demanding of working memory and therefore facilitate incidental vocabulary learning even more effectively.

The present study compares the effects of asynchronous reading-and-listening vs. synchronous reading-while-listening on incidental vocabulary learning. Twenty-seven participants, all adult English learners at B1 to C1 level, were assigned randomly to two comparison groups: synchronous and asynchronous. All participants read-and-listened to three stories, which contained 24 target words. After completing the stories, all participants immediately took surprise vocabulary tests on their knowledge of the target words, followed by delayed vocabulary tests one week later. Results showed that the asynchronous group gained statistically significantly higher scores on the immediate form-meaning test ($p=.04$). This suggests that the asynchronous approach was more effective in facilitating the incidental acquisition of at least some dimensions of vocabulary knowledge. However, there were no statistically significant differences between groups on the delayed vocabulary tests, so this advantage may not have been maintained over time. The dissertation concludes that asynchronous reading-and-listening appears to be an effective technique which deserves further study.

Contents

List of Tables 8
1. Introduction 9
2. Literature Review 12
2.1. Incidental Vocabulary Learning 12
2.2 Vocabulary knowledge 13
2.2.1 Dimensions of vocabulary knowledge 13
2.2.2 Measuring vocabulary knowledge 14
2.3. Incidental Vocabulary learning from reading 15
2.4. Reading-while-listening 17
2.4.1. Brown et al. (2008) 17
2.4.2. Webb and Chang (2012) 18
2.4.3. Malone (2018) 19
2.4.4. Teng (2018) 20
2.4.5. Vu and Peters (2020) 21
2.4.6. Chen (2021) 22
2.4.7. Summary 22
2.4.8. Gaps in the literature 23
2.5. Underlying mechanisms 24
2.5.1. Why is reading-while-listening more effective than reading only? 24
2.5.2. Cognitive load 25
2.5.3. The case for asynchronous reading-and-listening 27
2.6. Research Questions 28
2.6.1. Summary 28
2.6.2. Research questions 29
3. Methods 31
3.1. Study design 31
3.2. Participants 32
3.2.1 Recruitment 32
3.2.2. Description of Participants 33
3.2.3 Allocation to comparison groups. 33

3.3. Materials 35
3.3.1 Texts 35
3.3.2. Videos 37
3.3.3. Tests 40
3.3.4. The final product 42
3.4. Procedure 43
3.4.1. Session One 43
3.4.2. Session Two 45
3.5. Pilot 45
3.5.1. Procedure 45
3.5.2. Results of the pilot 46
3.6. Ethical considerations 47
4. Results 48
4.1. Data analysis 48
4.2. Participants and exclusions. 48
4.3. Comprehension tests 50
4.4. RQ1 50
4.4.1. Immediate form-recognition test 50
4.4.2. Immediate form-meaning test 51
4.5. RQ2: Delayed vocabulary tests 52
4.5.1. Delayed form-recognition test 52
4.5.2. Delayed form-meaning test 53
4.5.3. Within-group vocabulary retention 54
4.5.4. Between-group vocabulary retention 56
4.6. Summary 57
5. Discussion 59
5.1. RQ1: Immediate effects 59
5.1.1. Between-group differences on the immediate vocabulary test scores 59
5.1.2. Why might asynchronous reading-and-listening perform better than synchronous reading-while-listening? 59
5.1.3. Other observations about the immediate vocabulary test scores 61
5.2. RQ2: delayed effects 61
5.2.1. Between-group differences in the delayed vocabulary test scores 62

5.2.2. Vocabulary retention in the synchronous group 63
5.2.3. Vocabulary retention in the asynchronous group 64
5.2.4. Between-Group Differences in Vocabulary Retention 64
6. Conclusion 65
6.1. Research Questions re-examined 65
6.2. Pedagogical Implications 66
6.3. Limitations and recommendations for further research 67
Bibliography 69
Appendices 76
Appendix A: Texts 76
Appendix B: Videos 80
Appendix C: Comprehension Questions 81
Appendix D: Immediate form-meaning test 83
Appendix E: Immediate Form-Recognition Test 85
Appendix F: Delayed form-meaning test 87
Appendix G: Delayed form-recognition test 89
Appendix H: Webpages 91
Appendix I: Participant Information Sheet 92
Appendix J: Participant Consent form 96
Appendix K: Background Information Survey 98
Appendix L: Interview questions for pilot participants 100
Appendix M: Test results 101
Appendix N: CUREC approval email 102

List of Tables

Table 1: Characteristics of the Comparison Groups.....34
Table 2: participants whose data was retained for the comprehension and immediate form-recognition tests.....49
Table 3: participants whose data was retained for the immediate form-meaning test.....49
Table 4: participants whose data was retained for the delayed vocabulary tests.....50
Table 5: Form-recognition: Descriptive statistics.....51
Table 6: Form recognition: t-test.....51
Table 7: Form-meaning: Descriptive statistics.....51
Table 8: Form-meaning: t-tests.....52
Table 9: Form-recognition: descriptive statistics.....52
Table 10: Delayed form-meaning: descriptive statistics.....53
Table 11: Delayed form-meaning test: t-test.....53
Table 12: Immediate and delayed test scores: descriptive statistics.....54
Table 13: Immediate vs. delayed vocabulary test scores: paired t-tests.....55
Table 14: Between-group vocabulary retention (immediate score – delayed score): descriptive statistics.....56
Table 15: Between group vocabulary retention (immediate score-delayed score) on the form-recognition test: t-test.....57

1. Introduction

Learning vocabulary is an essential element of language acquisition: as Wilkins (1972, pp.110-111) notes, “without grammar, very little can be conveyed but without vocabulary, nothing can be conveyed”. Yet the challenge of acquiring sufficient vocabulary is considerable. Learners must know an estimated 8,000 word families to understand written texts, and 6,000 to 7,000 word families to comprehend spoken English (Nation, 2006). It is challenging for second language (L2) learners to acquire this vocabulary, because the time available to study vocabulary in classrooms is limited (Chen, 2021). Moreover, research suggests that only the 3,000 most frequent word families are commonly taught in classrooms (Schmitt & Schmitt, 2014).

Scholars have therefore suggested that an important role is played by incidental vocabulary learning: words may be acquired without deliberate attention as students read or listen to texts (Malone, 2018). The evidence suggests that vocabulary can be acquired incidentally by reading texts, but this is a slow and incremental process (Chen & Truscott, 2010; Waring & Takaki, 2003). As a result, there has been considerable interest in techniques which could increase the rate of incidental vocabulary acquisition from reading. These include underlining target words (Vu & Peters, 2021), providing glossaries (Abraham, 2008), and having teachers explain the meaning of unfamiliar words (Elley, 1989; Zhang & Graham, 2020). Such approaches have been shown to help language learners by enabling them to learn vocabulary at a faster rate.

One such technique is reading-while-listening: providing students with the print version of a text to follow while simultaneously listening to someone read it aloud. Several studies have found that reading-while-listening facilitates more incidental vocabulary learning than reading alone (Chen, 2021; Malone, 2018; Teng, 2018; Webb & Chang, 2012). From eye-tracking studies, it appears that learners who are reading-while-listening typically read ahead of the audio (Conklin et al., 2020). This behaviour may underpin the benefits of reading-while-listening, because seeing the form of the written text before hearing it helps learners segment and decode the aural input, enabling them to link the phonological and orthographic forms of new words (Conklin et al., 2020).

Yet cognitive load theorists argue that synchronous reading-while-listening also brings challenges (Diao and Sweller, 2007; Moussa-Inaty et al., 2012), because the task of focusing on and integrating two simultaneous streams of input places heavy and extraneous demands on working memory. In

turn, this is believed to reduce the capacity of working memory to focus on meaning and integrate new words into long term memory, ultimately potentially negatively affecting vocabulary learning.

In this dissertation, I aim to assess the effects of an alternative approach, asynchronous reading-and-listening, which may be more effective than synchronous reading-while-listening. Previous studies have focused on synchronous reading-while-listening, in which learners simultaneously read a text and listen to the audio. In contrast, asynchronous reading-and-listening involves reading a short section of the text, then listening to the same section, then reading the next short section, then listening to it, and continuing in this way throughout the text. This approach is expected to retain the advantages of synchronous reading-while-listening: learners can still read ahead of the audio, helping them decode the audio and link textual and aural forms. However, extraneous cognitive load should be reduced, because learners no longer need to focus on two simultaneous streams of input. This should free up working memory resources to carry out other important processes such as focusing on the meaning of new words and integrating them into long-term memory. As a result, asynchronous reading-and-listening is predicted to facilitate incidental vocabulary acquisition more effectively than synchronous reading-while listening.

To the best of my knowledge, this is the first study to formulate the concept of asynchronous reading-and-listening, and to assess its effects on incidental vocabulary acquisition. To address this gap in the literature, this study investigates whether more vocabulary is acquired through asynchronous reading-and-listening than synchronous reading-while-listening, under incidental learning conditions. To establish whether vocabulary gains were maintained over time, both immediate and delayed vocabulary post-tests were administered.

This testing regime also helps address a second gap in the literature. The majority of previous studies into synchronous reading-while-listening only tested participants' vocabulary gains immediately after the treatment: they did not include delayed vocabulary tests (Chen, 2021; Malone, 2018; Teng, 2018; Webb & Chang, 2012). As a result, while there is considerable evidence that the synchronous approach facilitates immediate vocabulary gains, little is known about the extent to which this vocabulary is retained over time. The present study helps to address this gap by carrying out both immediate and delayed vocabulary tests after synchronous reading-while-listening.

This dissertation is therefore guided by two research questions. First, what are the relative effects of synchronous reading-while-listening and asynchronous reading-and-listening on incidental vocabulary acquisition, as measured by vocabulary tests administered immediately after the treatment? Second, to what extent are any gains in vocabulary knowledge following synchronous reading-while-listening and asynchronous reading-and-listening maintained over time? The answers will help determine whether asynchronous reading-and-listening has the potential to increase incidental vocabulary acquisition for language learners.

2. Literature Review

This literature review aims to contextualise and justify my choice of research questions, and introduce concepts which will inform the methodological decisions made later in the dissertation. In section 2.1, the concept of incidental vocabulary learning is introduced and defined. Section 2.2 provides an overview of the different dimensions of vocabulary knowledge, and the various ways in which they can be tested; this information will inform the selection of instruments in the present study. Section 2.3 reviews the literature on incidental vocabulary acquisition from reading only; section 2.4 focuses on studies which compare the impact of reading only with reading-while-listening and identifies limitations and gaps in this literature. Section 2.5 discusses the mechanisms underlying reading-while-listening and draws out the implications of cognitive load theory to argue that asynchronous reading-and-listening could be a more effective approach. Finally, section 2.6 summarises the argument and presents the research questions which guide this dissertation.

2.1. Incidental Vocabulary Learning

There are several commonly used definitions of incidental vocabulary learning (Bisson et al., 2014; Brown et al., 2008; Webb, 2020). In this dissertation, I will adopt Webb's (2020, pp.227) definition of incidental vocabulary learning as "learning words as a by-product of meaning-focused activities or tasks". Thus, vocabulary is learnt incidentally from reading if the reading activity is focused mainly on understanding or engaging with the meaning of the text, rather than learning new words. This does not preclude students from focusing briefly on the meanings of unfamiliar words, so long as this was not the main goal of the activity (Webb, 2020).

This definition requires researchers to decide whether an activity is primarily focused on meaning. To operationalise this, many studies have followed Hulstijn's (2003) suggestion that incidental vocabulary learning takes place whenever participants are not warned in advance that a vocabulary test will follow the activity. This operationalization has the benefit of being precise and easy to control in experimental contexts, but I argue that it is insufficient, for two reasons. First, if participants are explicitly instructed to learn new vocabulary from a text, the learning conditional is by definition intentional rather than incidental, whether or not they were pre-warned about a vocabulary test, and it should be operationalised as such. Second, learning should also be regarded

as intentional if participants have clear reasons to expect a vocabulary post-test, even if they were not explicitly warned that one would take place (Alotaibi et al., 2022). For example, if participants read a text which is immediately followed by a vocabulary test, and are then given a second, similar text, they are likely to anticipate that this second text will also be followed by a test, even if they have not been explicitly told to expect this (Alotaibi et al., 2022). Given these objections, incidental vocabulary learning will be operationalised in this dissertation as taking place when three conditions are met: participants are not asked to learn vocabulary; nor are they warned in advance about a vocabulary post-test; nor do they have strong grounds to believe that such a vocabulary test will take place.

A related concept to incidental vocabulary acquisition is that of extensive reading: reading a large amount of interesting, comprehensible and self-selected material with a primary focus on understanding and enjoying what is read, rather than learning vocabulary or other language points (Nation, 2005). Vocabulary acquired through extensive reading is by definition learnt incidentally (Nation, 2005); therefore, this dissertation will draw on the literature on extensive reading when appropriate.

2.2 Vocabulary knowledge

This section reviews the literature on the different dimensions of vocabulary knowledge and discusses how some of these dimensions can be tested. This will inform the subsequent review of studies into incidental vocabulary learning and guide the selection of tests used in the present study. Particular attention is therefore paid to the types of tests which are commonly used in incidental vocabulary learning research.

2.2.1 Dimensions of vocabulary knowledge

Vocabulary knowledge is complex and multi-dimensional (Cheng & Matthews, 2018); various taxonomies of the elements of vocabulary knowledge have been proposed, but Nation's (2022) nine-part model is the most comprehensive and widely accepted. This model identifies nine *aspects* of vocabulary knowledge: three connected to form (knowledge of spoken form, written form, and word parts), three connected to meaning (knowledge of the form-meaning link, the concept and its

referents, and associations) and three connected to use (knowledge of grammatical functions, collocations, and constrains on use such as register). Each aspect can be further subdivided into *receptive* and *productive* knowledge: for example, for written form, receptive knowledge entails recognizing the word when it is written down, while productive knowledge entails the ability to write it. Moreover, knowledge of each aspect can vary in *strength*: some aspects may be known more precisely or fluently than others: for example, a learner may be able to spell a word correctly or only partially correctly, and they may be able to do so quickly and automatically, or only slowly and with conscious thought.

The evidence suggests that some dimensions of knowledge are typically learnt more quickly and easily than others (Nation, 2022). Phonological and orthographic form, along with form-meaning links, are generally the first aspects to be learnt (Nation, 2022), and receptive knowledge is generally acquired more quickly than productive knowledge (Ellis & Beaton, 1993). Incidental vocabulary learning studies have often focused on these dimensions in order to detect small amounts of knowledge in the early stages of learning a word (Webb & Chang, 2012).

2.2.2 Measuring vocabulary knowledge

Each dimension of vocabulary knowledge can be tested in different ways. A crucial distinction should be made between *recognition* tests, in which participants select the correct answer from a set of options, and *recall* tests, in which participants must produce the correct answer without such assistance (Nation, 2022). For example, in a recognition test of the link between form and meaning, participants might take a multiple choice test in which they see an L2 word and choose the correct definition of a word from a set of distractors; in a corresponding recall test, participants would see the L2 word and provide their own definition. Recognition tests are typically easier than recall tests (Nation, 2022), and can therefore be more suitable for detecting small amounts of knowledge.

Since there are many dimensions of vocabulary knowledge, each of which can be tested in different ways, scholars have unsurprisingly developed a multitude of different vocabulary tests, which assess constructs as diverse as grammar and collocation recognition (Teng, 2018) and online measures of constructs such as lexicalization (Chen 2021). This literature review will now focus on tests for two specific dimensions of vocabulary knowledge: receptive knowledge of written (orthographic) form,

and receptive knowledge of form-meaning links. These dimensions have been chosen because they are among the first to be learnt, making them suitable for detecting knowledge in the early stages of learning a word (Nation, 2022); as a result, they are widely used in the literature on incidental vocabulary acquisition from reading-while-listening (section 2.4) and are employed in the present study (section 3.3.3).

Receptive knowledge of orthographic form can be tested different ways, which vary in their degree of challenge. Easier tests require participants to view a list of target words and state whether they have encountered each word before (Webb & Chang, 2012); the target vocabulary can be presented alongside a list of distractors to check whether participants' self-reports are accurate and control for guessing (Malone, 2018). More difficult tests require participants to identify the correct spelling of a word from a list including similar distractors (Vu & Peters, 2020).

Receptive knowledge of the form-meaning link can also be tested in multiple ways. Meaning recall can be assessed by asking participants to provide a translation, synonym or definition for an L2 word (Teng, 2018). Meaning recognition can be assessed through multiple choice tests in which participants are shown the L2 word and must select a definition, translation or synonym from a list of distractors (Webb & Chang, 2012). Distractors from different semantic sets can be used to make the tests easier and thus more sensitive to small amounts of knowledge (Malone 2018). Participants can also simply be asked to self-report whether they know what the word means, without checking whether they are correct (Milton & Hopkins, 2006).

2.3. Incidental Vocabulary learning from reading

Incidental vocabulary acquisition from reading has been widely studied and often functions as the default approach against which other techniques, such as simultaneously reading-while-listening to a text, are compared (Malone, 2018; Webb & Chang, 2012). There is considerable evidence that reading can facilitate the incidental acquisition of vocabulary (Chen & Truscott, 2010; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Suk, 2017; Teng, 2014; Waring & Takaki, 2003; Webb et al., 2023). For example, Chen and Truscott (2010) investigated the incidental learning of ten rare English words. Their participants, 72 intermediate English as a Foreign Language (EFL) students at a Taiwanese university, were randomly assigned to read one of three different sets of

texts, in which the target words were repeated one, three, or seven times. After reading the texts, all participants took a surprise vocabulary test on seven dimensions of vocabulary knowledge, followed by an identical delayed test two weeks later. Results showed that participants gained both orthographic and semantic word knowledge. Words that were encountered more frequently were statistically significantly more likely to be learnt ($p<0.05$). However, some dimensions of knowledge, including knowledge of meaning, were acquired rather slowly, with even seven exposures being insufficient to learn the meaning of more challenging words. Mean scores on the delayed tests were typically lower than on the immediate tests, suggesting that some knowledge had been forgotten.

These findings were largely corroborated by Teng (2014). Twenty Chinese college students read a graded reader in which 30 words had been replaced with pseudowords; each pseudoword was repeated between one and 20 times in the text. The participants took a vocabulary post-test which measured their receptive knowledge of orthographic form, meaning, and use. Again, results showed that vocabulary was learnt from the graded reader, but a large number of repetitions was typically required to gain vocabulary knowledge: a word had to be repeated 14 times to give learners a 50% chance of remembering its meaning and usage. Different dimensions of knowledge were gained at different rates: form was learnt most easily, followed by meaning, followed by usage. No delayed test was administered, so it is unclear to what extent vocabulary gains were retained over time.

Both studies suggest that suggest that some dimensions of knowledge, such as receptive knowledge of form, can be acquired more quickly than others. This conclusion has been widely corroborated by other studies (Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Waring & Takaki, 2003). More broadly, Chen and Truscott's (2010) and Teng's (2014) results suggest that vocabulary can be learnt incidentally from reading, but that this is a somewhat slow process and not all unfamiliar words are learnt. This conclusion is supported by Webb et al.'s (2023) systematic review, which analysed 24 studies into incidental vocabulary learning from texts; all studies included in the systematic review used control groups. In a meta-analysis of the results of these studies, Webb et al. (2023) found that vocabulary was learnt incidentally from reading texts, but the proportion of target words learnt was typically rather low: participants in most studies acquired fewer than 20% of the target words.

The extent to which vocabulary is forgotten between immediate and delayed post-tests is somewhat unclear. Chen and Truscott (2010) found that delayed vocabulary test scores were

typically lower than immediate post-test scores, and similar results have obtained in some other studies (Brown et al., 2008; Waring & Takaki, 2003). However, Webb et al.'s (2023) meta-analysis found that while scores on some dimensions of vocabulary knowledge fell from immediate to delayed tests, the scores on other dimensions actually rose. Webb et al. (2023) speculate that this somewhat surprising result could be due to practice effects: the process of taking the immediate post-test could itself have helped participants deepen their vocabulary knowledge, enabling them to perform better on the delayed post-test. Given the variation in results between studies, more research is needed into the retention of incidentally acquired vocabulary over time.

In summary, it appears that vocabulary can be acquired incidentally from reading texts, but this is a relatively slow and inefficient process. Some dimensions of vocabulary knowledge, such as receptive knowledge of form, are usually learnt more quickly than others. It is unclear to what extent vocabulary that has been learnt incidentally is subsequently forgotten.

2.4. Reading-while-listening

Scholars have investigated whether the process of reading can be enhanced to improve the rate of incidental vocabulary acquisition. One such enhancement is reading-while-listening to a text: learners are given a text to read and simultaneously listen to an audio recording (or live reading) of it. This process is sometimes referred to as assisted reading (Webb & Chang, 2012) or aural enhancement (Malone, 2018). Studies have compared the impact of reading-while-listening and reading-only on helping students understand aural texts (Chang, 2009), improving reading rate and reading comprehension skills (Chang & Millet, 2015) learning collocations (Alotaibi et al., 2022; Vu & Peters, 2022; Webb & Chang, 2022), and learning vocabulary intentionally (Diao and Sweller, 2007; Moussa-Inaty, 2012). Six studies have attempted to compare the impact of reading only and reading-while-listening on the incidental acquisition of individual words. These studies are reviewed below.

2.4.1. Brown et al. (2008)

Brown et al. (2008) used three graded readers, replacing 28 words from each with pseudo-words. The participants were 35 Japanese university students, all pre-intermediate or intermediate, in

three intact classes which were assigned randomly to one of three conditions: reading only, listening only, and reading-while-listening. Participants were told that the purpose of the activity was to learn new vocabulary. Each group read and/or listened to each graded reader and then took an immediate vocabulary post-test, followed by two delayed post-tests, one week later and three months later. In each test, participants first attempted to translate each pseudo-word into Japanese and then select their meaning from a multiple choice. The test for the listening only group used aural prompts while the tests for the other groups used written prompts.

Results of the immediate post-test showed that all groups gained vocabulary knowledge according to both measures, with lower scores on the translation test. There was no statistically significant difference between the reading only and reading-while-listening groups, but both statistically significantly outperformed the listening only group. Multiple choice scores were largely maintained in the two delayed post-tests, with no statistically significant differences between immediate and delayed test results. However, the scores on the translation tests displayed a statistically significant decline from the immediate to delayed post-tests. In the reading only and reading-while-listening groups, words that were repeated more frequently were statistically significantly more likely to be learnt ($p<0.01$).

This study has several limitations. First, the sample size was small, with 35 students split into three groups, limiting its ability to identify statistically significant relationships. Second, the groups did not use identical tests: the listening group had aural prompts which could have been more challenging than the written prompts used by the other groups (Chang & Read, 2013). This undermines the conclusion that listening only was less effective for vocabulary learning. Third, although the authors describe the vocabulary learning as incidental, it should really be seen as intentional, because participants were asked to learn vocabulary and told that this was the purpose of the activity (Malone, 2018).

2.4.2. Webb and Chang (2012)

Webb and Chang (2012) investigated 82 beginner EFL students at a secondary school in Taiwan, aged 15 to 16. They were split into two groups, one reading only and one reading-while-listening, which were found to have no statistically significant differences on vocabulary pre-test scores. Both

groups read fourteen stories, two per week for seven weeks, reading each story at least twice, then repeated the process with another 14 stories over another seven-week period. Participants were allowed to use dictionaries if they wanted but were told that the main goal was to understand and enjoy the stories. One hundred target words were selected, 50 from each wave of 14 stories. To test word knowledge, participants self-reported whether they recognized a word and attempted to provide a translation. Participants were pre-tested before each seven-week period; the same test was used as a post-test one week after each seven-week period. Results showed that both groups made statistically significant vocabulary gains but the reading-while-listening group statistically significantly outperformed the reading-only group ($p=0.003$ over the first seven-week period and $p=0.004$ over the second one).

This study improves on Brown et al. (2008) by using a larger sample, employing identical written tests for both groups, and not asking students to learn vocabulary; it also has strong ecological validity by using multiple texts over a long time period. The use of a pre-test increases confidence that the words were not previously known to participants, but brings two disadvantages. First, it could have prompted students to learn vocabulary intentionally or expect a vocabulary post-test, which would render the learning condition intentional rather than incidental (Nation & Webb, 2011). Second, vocabulary could have been learnt from the pre-test as well as, or instead of, from the texts: thus, the form recognition knowledge detected in the post-tests could have been caused by the pre-test or by an interaction between taking the pre-test and reading the texts (Nation & Webb, 2011). There are also other limitations: no delayed post-test was used, so we do not know whether the vocabulary gains were maintained over time (Kremmel, 2020), and information about how participants were assigned to groups is lacking.

2.4.3. Malone (2018)

Malone (2018) used four texts, each containing eight real but very infrequent English target words. Two versions of each text were created, in which the target words were repeated either two or four times. The participants, 80 students of intermediate English at US universities, were randomly assigned to four groups: reading-only vs reading-while-listening, and using texts with two or four repetitions of each target word. Participants read the texts on a timed PowerPoint presentation and were asked to focus on answering comprehension questions. The reading-while-listening group

simultaneously read the text and listened to an audio recording; each line of written text on the PowerPoint vanished after they had heard it. For the reading only group, the lines of written text also vanished from their screens at the same rate. A surprise written post-test assessed form recognition (participants selected the target words from a list of distractors) and form-meaning links (participants selected the meaning of the target words from a list of distractors). Results showed that four encounters with a word prompted more learning than two encounters, in both reading-while-listening and reading-only. On the texts with two repetitions of each word, reading-while-listening outperformed reading-only on both form recognition and form-meaning; at four repetitions, reading-while-listening outperformed reading-only on form-meaning tests but there was no statistically significant difference on form recognition.

Vocabulary learning in this study was truly incidental: participants were not asked to learn words and had no reason to expect a post-test. However, this came at the cost of a reduction in ecological validity: the reading process was highly controlled, with lines of text appearing and vanishing from the screen as they were read or heard. Again, no delayed post-test was used, making it impossible to determine whether the effects persisted.

2.4.4. Teng (2018)

Teng's (2018) participants, 60 EFL students at a Chinese university, were randomly assigned to two groups: reading only and reading-while-listening. Both groups spent 70 minutes reading a graded reader in which 24 target words had been replaced with pseudo-words; the reading-while-listening group simultaneously heard the story read aloud; all participants were told that the goal was to understand the text. A surprise written post-test assessed four dimensions of word knowledge: form recognition, grammar recognition, meaning recall and recognition of collocation. For both reading-while-listening and reading-only, results were statistically significantly higher ($p<0.05$) for form recognition than for grammar recognition, which in turn was statistically significantly higher ($p<0.05$) than both meaning recall and recognition of collocation. Target words that occurred more frequently in the text gained statistically significantly higher scores ($p<0.05$) on both form recognition and grammar recognition. Finally, reading-while-listening outperformed reading-only on all four tests ($p<0.05$). This experiment involved truly incidental learning and had higher ecological

validity than Malone's (2018) study because the participants read the text in a more natural way. Again, however, there was no delayed post-test.

2.4.5. Vu and Peters (2020)

Vu and Peters (2020) carried out two experiments. The first used four texts, each containing four items of target vocabulary. The participants, 32 beginner EFL students at a Vietnamese university, were assigned to three groups: reading only, reading-while-listening, and reading with textual enhancement (meaning that target vocabulary was underlined). Each group read one text per week. For each text, participants were first pre-tested on the target vocabulary; one week later, they read the text and did an immediate post-test; one week afterwards, they did a delayed post-test. The whole experiment ran over a six week period: in week one, participants did a pre-test for the first test; in week two, they read the first text, did an immediate vocabulary post-test on it, and also completed the pre-test for the second text, and so on. Tests assessed form recognition (participants chose the correct spelling), form recall (participants translated a Vietnamese word to English) and meaning recall (participants translated an English word to Vietnamese). The second experiment followed the same format, but with different texts (each containing between two and four target words), and different participants: 28 intermediate Vietnamese university students. The results of both experiments showed that vocabulary was learnt in all three groups, and the textual enhancement group outperformed the others; there was no statistically significant difference between reading-while-listening and reading-only on any measures. Scores on the delayed vocabulary post-tests were slightly lower than on the equivalent immediate vocabulary post-tests, but the authors did not check whether any of these declines were statistically significant.

The study has several limitations. First, as the authors acknowledge, both experiments had small sample sizes due to participants dropping out; this drop-out rate could also have affected the validity of the statistical analyses performed. Moreover, each text contained only two to four target words (Vu & Peters, 2020), further reducing the study's statistical power. Second, the vocabulary tests were challenging, which may have limited their ability to detect small gains in knowledge: even the form recognition test required the students to select the correct spelling of the word rather than simply indicate whether they had seen the word before. Third, the learning condition was not truly incidental according to the definition used in this dissertation (section 2.1). This is

because the participants had strong grounds to expect vocabulary post-tests during the majority of the experiment: after taking a vocabulary test immediately after reading the first text in week one, participants were surely likely to expect analogous tests in subsequent weeks (Alotaibi et al., 2022).

2.4.6. Chen (2021)

Chen (2021) used the same texts as Malone (2018) but replaced the target vocabulary with pseudowords. The participants, 50 high intermediate ESL students at a US university, were randomly assigned to reading-only or reading-while-listening groups. As in Malone's (2018) study, the written text was presented on a series of PowerPoint slides on which lines of text appeared and disappeared as they were heard or read. Participants then took a surprise written post-test using the same measures of form recognition and form-meaning links as Malone (2018); results showed that reading-while-listening statistically significantly outperformed reading only with a medium effect size on both form-recognition (p=.02, d=.70) and form-meaning tests (p=.01, d=.74). Two online measures of vocabulary learning were also applied, but neither detected learning in either group.

This study was closely based on Malone (2018). It had similar results and similar strengths and limitations: learning was truly incidental, but at the cost of a reduction in ecological validity, and there was no delayed post-test.

2.4.7. Summary

Three conclusions can be drawn from these studies. First, reading-while-listening appears to facilitate more incidental vocabulary learning than reading-only, though more research is needed to strengthen this conclusion as not all studies agree. Of the six studies, four found that reading-while-listening statistically significantly outperformed reading-only, while two found no significant differences. Notably, the two studies that detected no advantage for reading-while-listening used smaller sample sizes and did not involve truly incidental learning conditions according to the definition adopted in section 2.1: Brown et al. (2008) because participants were asked to learn vocabulary from the texts, and Vu and Peters (2020) because participants could reasonably have expected vocabulary post-tests. In contrast, the four studies which found that reading-while-

listening outperformed reading-only all had larger samples and more closely approximate incidental learning. We can therefore tentatively conclude that reading-while-listening appears to facilitate more vocabulary learning than reading-only under genuinely incidental learning conditions. The advantage of reading-while-listening seems to hold for participants at beginner (Webb & Chang, 2012), intermediate (Malone, 2018) and high intermediate (Chen, 2021) proficiency levels.

Second, more exposures to a word during reading-while-listening appears to prompt more incidental vocabulary learning: the three studies which investigated this all found a statistically significant relationship between number of exposures and some (Teng, 2018) or all (Brown et al., 2008; Malone, 2018) measures of vocabulary knowledge. However, estimates of the number of exposures required vary widely. Brown et al. (2008) found that even after encountering a word fifteen to twenty times, participants only had a 26% chance of being able to translate it into their L1, but other studies that used more sensitive measures found effects at much lower frequencies: for example, Malone (2018) detected learning from two exposures and Teng (2018) at one to two exposures. The highest scores were gained on form recognition tests which required participants to either self-report that they had seen the word before (Webb & Chang, 2012), or select the word from a list of distractors (Brown et al., 2008; Chen, 2021; Malone, 2018; Teng, 2018). Thus, the third conclusion to be drawn from these studies is that test type has a powerful effect, with sensitive tests able to detect learning from as little as one exposure to a word.

2.4.8. Gaps in the literature

The above review of previous reading-while-listening studies reveals three priorities for future research. First, since not all studies found that reading-while-listening outperformed reading-only, more studies could compare the two conditions to strengthen our confidence in this conclusion. Second, such studies should create genuinely incidental learning conditions in which participants are not asked to learn vocabulary and have no reason to anticipate a vocabulary post-test.

Third, and most importantly, more studies into reading-while-listening should include delayed vocabulary post-tests because there is a particular lack of evidence about the extent to which vocabulary acquired incidentally from reading-while-listening is retained over time. Only two of the six existing studies used a delayed post-test: Brown et al. (2008) and Vu and Peters (2020).

Unfortunately, as discussed in section 2.4.7, neither study created truly incidental learning conditions. As a result, to the best of my knowledge, no studies have successfully investigated the rate of retention or decay of vocabulary learnt from reading-while listening under genuinely incidental learning conditions.

2.5. Underlying mechanisms

This section reviews the literature on the psychological mechanisms involved in reading-while-listening. Section 2.5.1 evaluates possible explanations for the effectiveness of reading-while-listening. Section 2.5.2 presents the argument that, although reading-while-listening is an effective approach, it may not be optimal because it imposes heavy demands on working memory, which may limit learners' ability to add new words to long-term memory. It is therefore suggested in section 2.5.3 that an alternative approach, asynchronous reading-and-listening, may be more effective.

2.5.1. Why is reading-while-listening more effective than reading only?

Several explanations have been proposed for the apparent advantage of reading-while-listening over reading-only in facilitating incidental vocabulary acquisition. These interpretations broadly fall into two camps. One set of explanations emphasizes the role that the aural input plays in helping learners comprehend the orthographic text: thus, Teng (2018) suggests that the audio helps learners group the written words into meaningful chunks. In contrast, another other set of explanations argues that the written text helps learners understand the audio, perhaps by enabling them to decode the stream of aural input into distinct words (Conklin et al., 2020). While both sets of explanations are possible, the argument that the written text facilitates comprehension of the audio is more compelling because it is compatible with the fact that most learners have more advanced reading skills than listening skills (Hirai, 1999): this suggests that learners would be more likely to use their stronger reading skills to support their weaker listening skills, rather than the other way round.

A study by Conklin et al. (2020) also supports the claim that learners use the written text to decode the audio. The researchers used eye tracking technology to observe the online behaviour of 28

advanced L2 speakers of English who were asked to read and listen simultaneously to two stories. Results showed that the participants' reading was often slightly misaligned with the audio. Participants read along with the audio only 33% of the time; in contrast, they read ahead 53% of the time. The results of this one study cannot be assumed to apply to all learners, especially since it involved highly proficient participants who may behave differently from less advanced learners. Nevertheless, this study raises the intriguing possibility that the advantages of reading-while-listening may be caused by reading ahead of the audio. Reading ahead could help learners identify word boundaries and thereby make sense of the upcoming aural input, subsequently enabling them to form links between the phonological and orthographic forms of the word (Conklin et al., 2020).

In summary, there is insufficient evidence to be certain about the reason why reading-while-listening appears to facilitate more vocabulary gains than reading-only. However, the balance of the evidence suggests that reading-while-listening is effective because learners typically read the orthographic text first and use this to help them decode the aural input, which in turn enables them to form mental links between the written and aural forms of new vocabulary.

2.5.2. Cognitive load

Despite the advantages of reading-while-listening, cognitive load theory suggests that it may also present learners with some challenges. This section introduces cognitive load theory and then relates it to reading-while-listening.

Cognitive load theory is based on the following model of human cognition (Moussa-Inaty et al., 2012). For new skills or information to be learnt, they must be added to long term memory. Long term memory has the capacity to store very large quantities of knowledge. However, for knowledge to be added to long term memory, it must be processed by working memory, which has a very limited capacity. The capacity of working memory is therefore a key constraint on the rate at which new knowledge can be acquired. This implies that pedagogical activities should direct working memory towards processes which help transfer knowledge to long-term memory. Such processes generate *germane* cognitive load, defined as demands on working memory which help build long term memory. In contrast, pedagogical activities should avoid processes which place demands on

working memory without helping to transfer the target knowledge to long term memory; the unhelpful cognitive load generated by such activities is defined as *extraneous* cognitive load.

One implication of cognitive load theory is the redundancy principle (Moussa-Inaty et al., 2007). This states that simultaneously presenting the same information in two ways can be detrimental for learning because the process of paying attention to both sources of input, and integrating them, takes up working memory resources. This generates extraneous cognitive load, leaving less working memory available to carry out processes which help commit the new information to long term memory.

The redundancy principle has implications for reading-while-listening, which requires learners to read and listen to the same information simultaneously, and integrate the two streams of input. Cognitive load theorists (Diao & Sweller, 2007; Moussa-Inaty et al., 2012) argue that this is likely to place a heavy demand on working memory. In turn, this leaves less working memory available to infer the meaning of unfamiliar words from context (Nation, 2022) and integrate these words with existing schemata (Moussa-Inaty et al., 2012).

These demands on working memory may be further increased if learners read ahead of the audio, which Conklin et al.'s (2020) eye-tracking study suggests may be a common occurrence. If the aural and orthographic input is misaligned, learners must process two different words simultaneously, which could limit comprehension (Conklin et al., 2020) and exacerbate processing demands on working memory.

The discussion thus far has been theoretical, but there is some evidence to support the claim that reading-while-listening in L2 increases cognitive load. In Daio and Sweller's (2007) study, 60 participants from a Chinese university were randomly assigned to one of two conditions: reading-only and reading-while-listening. The participants were given two texts and asked to give a subjective rating of how difficult they were to understand. One group of participants simultaneously read and listened to the two texts, while the other group only read the texts, without any audio. Results showed that participants in the reading-while-listening condition found one of the texts statistically significantly more difficult to understand compared to participants in the reading only condition, though there were no statistically significant between-group differences for the other text. This suggests that reading-while-listening was more cognitively demanding than reading-only, as predicted by the redundancy principle. These results are also corroborated by qualitative data

from Vu and Peters' (2020) study (section 2.4.5): several participants in the reading-while-listening group commented that they found the simultaneous streams of aural and written input distracting, making it harder for them to pay attention to new words.

This evidence base must be interpreted with caution because it consists of only two studies, one of which (Diao and Sweller, 2007) investigated intentional rather than incidental vocabulary learning. Nevertheless, we can tentatively conclude that the evidence supports the prediction that reading-while-listening places more demands on working memory than reading-only. This may reduce the capacity of working memory to focus on other important processes such as inferring meaning from context, thereby limiting vocabulary learning.

2.5.3. The case for asynchronous reading-and-listening

The argument made so far will now be summarised. Reading-while-listening appears to facilitate more incidental vocabulary acquisition than reading-only. The most probable explanation for this is that learners who are reading-while-listening typically read ahead of the audio, use the text to decode the aural input, and subsequently form mental links between the aural and orthographic forms of new vocabulary. However, although reading-while-listening is an effective approach, there is good reason to believe that it may not be optimal, because the process of simultaneously focusing on both aural and textual input, which are often misaligned, requires a lot of working memory. This means that there is less working memory available to focus on other processes which are important for learning new vocabulary, such as inferring the meaning of unfamiliar words from context and integrating new vocabulary with existing schemata.

This suggests that an alternative process, asynchronous reading-and-listening, may be more effective. Asynchronous reading-and-listening involves first reading a small section of text, then listening to it, then reading another small section, then listening to it, and proceeding in this way throughout the entire text. This contrasts with synchronous reading-while-listening, the approach investigated by the studies reviewed in section 2.4, in which participants read the text and listen to the audio at exactly the same time. The asynchronous approach would be expected to bring the same benefits as synchronous reading-while-listening, because learners can still read ahead of the audio and use the written text to decode the aural input. Yet the asynchronous approach should

impose fewer extraneous demands on working memory, because learners do not need to focus on two input streams simultaneously. This suggests that asynchronous reading-and-listening would free up working memory to carry out other processes which promote vocabulary learning and should therefore facilitate more incidental vocabulary learning than synchronous reading-while-listening.

To the best of my knowledge, the present study is the first to formulate the concept of asynchronous reading-and-listening, and its impact on incidental vocabulary acquisition has never previously been tested. There is therefore no evidence to support or contest the theoretical argument presented above. The present study seeks to contribute to addressing this gap in the literature by comparing the impact of synchronous reading-while-listening and asynchronous reading-and-listening on incidental vocabulary acquisition.

2.6. Research Questions

2.6.1. Summary

The key points made in this literature review can now be drawn together. To learn a language, large amounts of vocabulary must be acquired. This is an incremental process which involves developing multiple dimensions of knowledge. Proficient language users appear to acquire much of their vocabulary incidentally, through activities such as reading. However, incidental vocabulary learning from reading is a slow process. Simultaneously reading-while-listening to a text appears to increase the rate of incidental vocabulary acquisition, probably because it enables learners to read ahead of the audio, which helps them decode the aural input and link the orthographic and phonological forms of new words. However, it is unclear how well the vocabulary learnt from reading-while-listening is retained, because previous studies have not typically included delayed tests. Moreover, while reading-while-listening seems to be an effective approach, there are reasons to believe that it may not be optimal, because the task of paying attention to and integrating written and aural input simultaneously rather than sequentially may impose extraneous cognitive load. This suggests that asynchronous reading-and-listening could facilitate incidental vocabulary acquisition more effectively by bringing the benefits of synchronous reading-while-listening while reducing extraneous cognitive load.

This dissertation aims to test this hypothesis by investigating whether asynchronous reading-and-listening facilitates more incidental vocabulary acquisition than synchronous reading-while-listening, as measured by immediate and delayed vocabulary tests. While investigating this, the present study will also help to fill a second gap in the literature by generating evidence on the extent to which vocabulary learnt from synchronous reading-while-listening is retained over time.

2.6.2. Research questions

In the light of this discussion, the following research questions can be formulated.

RQ1: *What are the relative effects of synchronous reading-while-listening and asynchronous reading-and-listening on incidental vocabulary learning, as measured by vocabulary tests administered immediately after the treatment?*

The asynchronous approach is expected to be more effective than the synchronous approach (section 2.5.3). Therefore, the following hypotheses have been formulated:

> *H_1: More vocabulary will be learnt from asynchronous reading-and-listening than from synchronous reading-while-listening, under incidental learning conditions, as measured by immediate post-tests*
>
> *H_0: No more vocabulary will be learnt from asynchronous reading-and-listening than from synchronous reading-while-listening, under incidental learning conditions, as measured by immediate post-tests*

RQ2: *To what extent are any gains in vocabulary knowledge following synchronous reading-while-listening and asynchronous reading-and-listening maintained over time, as measured by delayed vocabulary tests?*

Again, since the asynchronous approach is expected to be more effective than the synchronous approach (section 2.5.3), the following hypotheses have been formulated:

H_1: More vocabulary will be retained following asynchronous reading-and-listening than from synchronous reading-while-listening, under incidental learning conditions, as measured by delayed post-tests

H_0: No more vocabulary will be retained following asynchronous reading-and-listening than from synchronous reading-while-listening, under incidental learning conditions, as measured by delayed post-tests

3. Methods

3.1. Study design

The present study adopted a two arm multi-centre parallel group randomised superiority trial design, with stratified random allocation of participants. A two-armed superiority design was appropriate because the study aimed to compare the performance of two approaches, synchronous reading-while-listening and asynchronous reading-and-listening, with the goal of establishing whether either of these approaches is superior to the other. A multi-centre approach was used to increase the number of participants in the study (section 3.2.1). Random allocation of participants to the comparison groups was employed because this is the best way of allocating unknown or unmeasurable variation fairly between groups, ensuring that any differences between the comparison groups at baseline are due only to chance (Torgerson & Torgerson, 2008). Specifically, stratified random allocation was used to ensure that participants recruited from different centres, and with different proficiency levels, were fairly distributed between the groups (section 3.2.3).

Twenty-seven English language learners, all at a B1 to C1 level, were recruited from five centres and randomly assigned to one of two comparison groups: asynchronous reading-and-listening, and synchronous reading-while listening. All participants read and listened to three stories, which contained a total of 24 target words believed to be previously unknown to the participants. In the synchronous group, participants simultaneously read and listened to each story. In the asynchronous group, for each story, participants first read a short section, then listened to that section read aloud, then read the next short section, then listened to that section, and continued until they had completed the story. After completing the stories, all participants immediately took surprise vocabulary post-tests on their knowledge of the target words, followed by delayed vocabulary post-tests one week later.

3.2. Participants

3.2.1 Recruitment

The participants were recruited through five language teachers known personally to the researcher. Each teacher worked at a different centre: one worked at an Iraqi university, one worked at a Korean university, one worked at a private language school in the UK, one was a self-employed English teacher working in the UK, and one was a self-employed English teacher working in Egypt. Each teacher gave their students some basic information about the study and then passed on the details of any interested students to the researcher. The researcher then provided the potential participants with more comprehensive information and obtained their informed consent (section 3.6). This multi-centre approach, involving five teachers, was used to ensure that a sufficiently large number of participants were recruited.

Participants were accepted based on two inclusion criteria: they were aged 18 or over, and they had been assessed by their teachers or language teaching institutions to be at a B1, B2 or C1 level. Accepting participants at multiple levels enabled the recruitment of a larger number of participants, while making the group less homogenous. This was judged to be an acceptable trade-off because stratified randomised allocation was used to ensure that participants of different proficiency levels were evenly distributed between the two comparison groups (section 3.2.3). Moreover, the materials were judged to be suitable for students ranging from a low B1 to high C1 level: as explained in section 3.3.1, the texts were simple enough to be understood by B1 level participants, while the target vocabulary was so advanced that it was expected to be mostly unfamiliar even to C1 level participants.

Placement tests were not administered to check these levels, to avoid imposing an undue burden on participants which could have negatively affected recruitment. Instead, the judgement of the participants' teachers or language teaching institutions were relied upon. All the teachers were qualified, highly experienced, and aware of the CEFR descriptors. Nevertheless, relying on the judgement of different teachers and institutions to assess the students' level raised the risk of there being systematic differences in proficiency between participants from different institutions. To mitigate this risk, stratified randomised allocation was used to compensate for any such systematic variation (section 3.2.3).

Recruitment was paused once 28 participants had been recruited. It would have been preferable to recruit a larger number if more time had been available; however, given the time constraints faced, 28 participants was judged to be an acceptable number as it would enable approximately 14 participants to be assigned to each comparison group, a greater number than that achieved in several previous reading-while-listening studies (Brown et al., 2008; Vu & Peters, 2020).

3.2.2. Description of Participants

Initially, 28 participants were recruited. However, one participant followed the procedure incorrectly and his results were excluded (section 4.2), leaving a final total of 27 participants.

Of these 27 participants, 16 were female, 10 were male, and one preferred not to say. The majority identified their L1 as Arabic (n=22) or Egyptian Arabic (n=2); other L1s were Korean (n=1), Brazilian Portuguese (n=1) and Russian (n=1). Ages ranged from 18 to 51 years, with a mean age of 26.9. Participants had been studying English for between four and 45 years, with a mean of 14.6 years (standard deviation = 9.72). In terms of the highest level of education completed, all participants had completed high school; 18 were studying for an undergraduate degree but had not yet completed it; four had completed an undergraduate degree only; and five had also completed a postgraduate degree. Six participants were at a B1 proficiency level; 16 were B2; four were C1; and one was assessed by their teacher as B1/B2.

3.2.3 Allocation to comparison groups.

Participants were allocated to one of two comparison groups: synchronous reading-while listening and asynchronous reading-and-listening. Stratified random allocation was used to account for the fact that participants' proficiency levels varied from B1 to C1, as well as possibility that participants recruited from different language learning institutions may have differed systematically on variables related to their learning institution and location, which could have affected their incidental acquisition of vocabulary. Given the relatively small sample size of 27, there was a risk that simple randomised allocation could have resulted in the concentration of participants from certain centres, or at certain proficiency levels, in just one of the two comparison groups. Stratified randomised allocation can be used to solve this problem by ensuring that approximately equal numbers of

participants from each centre are assigned to each comparison group (Torgerson & Torgerson, 2008).

The allocation was carried out as follows. Participants were divided into strata according to their CEFR level and the centre they had been recruited from. For example, the 17 participants recruited from the Iraqi university were divided into two strata: one at B1 level, containing four participants, and one at B2 level, containing 13 participants. In total, eight strata were generated through this process. Participants within each stratum were then randomly assigned to the synchronous and asynchronous groups using the online list randomiser at random.org, which generates truly random numbers using atmospheric noise (Haahr, 2023): a list of the names of participants in each stratum was entered into the list randomiser which reordered them randomly. The top half of the list was then assigned to the asynchronous group and the bottom half to the synchronous group; when there was an odd number of participants in a stratum, the participant in the middle was randomly assigned to either the synchronous or asynchronous group by the flip of a coin. For strata containing only one participant, the participant was likewise allocated by a coin flip. This process resulted in the allocation of 14 participants to the asynchronous group and 13 participants to the synchronous group. These characteristics of these two groups are set out in table 1, below:

Table 1: Characteristics of the Comparison Groups

	Synchronous Group	**Asynchronous Group**
Number of participants	13	14
Mean age	26.46	27.43
Mean number of years spent studying English	13.42	15.64
Gender	Female: 10 Male: 3	Female: 6 Male: 7 Prefer not to say: 1
Proficiency level	B1: 3 B2: 8 C1: 2	B1: 3 B1/B2*: 1 B2: 8

		C1: 2
Self-reported L1	Arabic: 11 Egyptian Arabic: 1 Russian: 1	Arabic: 11 Egyptian Arabic: 1 Korean: 1 Portuguese (Brazil): 1
Highest level of education completed	High School: 9 Undergraduate degree: 2 Postgraduate degree: 2	High School: 9 Undergraduate degree: 2 Postgraduate degree: 3
Number of participants recruited from each centre	Iraqi university: 9 Private teacher in Egypt: 3 Private teacher in UK: 1	Iraqi university: 8 Private teacher in Egypt: 4 Korean university: 1 UK language school: 1

*The teacher believed this student to be either B1 or B2 but could not be more specific

3.3. Materials

This study used modified versions of texts and tests created by Malone (2018). These materials were judged to be well suited to the present study after some modifications (sections 3.3.1 and 3.3.3). Moreover, they had the advantage of having already been used successfully in two studies: both Malone (2018) and Chen (2021) used these materials, and both found that the tests detected knowledge of some but not all target words after participants had simultaneously read and listened to the texts. This suggested that the materials would be well calibrated to detect variation between comparison groups in the present study.

3.3.1 Texts

Three texts were used (appendix A). These were based on texts created by Malone (2018) and had been shortened slightly, as explained below. These texts were chosen because they were judged to meet the requirements of the present study, and because they had already been used successfully in two previous studies (Malone, 2018; Chen, 2021). Each text told a short fictional story, which were judged to be accessible to a wide range of students because they did not require specific

background knowledge; participants in the pilot confirmed that they had no difficulty understanding the stories.

Malone's (2018) study used four texts, but only three of these were used in the present study to ensure that the treatment could be completed by participants in a reasonable time. This change in methodology was necessary because the asynchronous treatment delivered in the present study took twice as long as the synchronous approach used by Malone; it was judged to be important to limit the total treatment time to around 80 minutes in order to maintain participants' participation and focus.

Each text was approximately 600 words long. The original texts made by Malone were between 694 and 773 words long, but they were shortened slightly to further reduce the treatment time. The texts were shortened by paraphrasing or deleting small portions of the texts while preserving the central narrative of each story.

There were 24 target words across the three texts, eight in each text. All target words were real but extremely rare English words, being outside the 25,000 most frequent lemmas according to the Corpus of Contemporary American English (COCA) (Davies, 2008). The present study followed Malone (2018) in using real target words rather than pseudowords, for ethical reasons: by learning real words, participants could gain something useful from their involvement in the experiment. Moreover, the use of real words makes the study more realistic: learners may react differently to pseudowords than they do to real words, possibly because pseudowords are more salient (Uchihara et al., 2019). The use of real words brings the risk that some participants may already know some of the target words. However, even if some participants did know some of the target words, this would not affect the utility of the study, so long as enough of the target words were unknown to detect variation in vocabulary acquisition between the comparison groups. Participants were judged to be unlikely to know a large number of the target words prior to joining the study because even C1 level students have been estimated to have an average vocabulary size of less than 5,000 lemmas (Milton & Alexiou, 2009), while the target words were outside the top 25,000 most frequent lemmas. This belief was supported by evidence from the pilot: two of the three pilot participants confirmed that they had not encountered any of the target words before, while the third participant had encountered only one of the 24 words previously.

Each target word appeared twice in the text; as a result, the target words constituted between 2.6% and 2.7% of the running words in each text[1]. The other words in the texts were relatively common: at least 97% of the running words in each text was within the 3,000 most frequent lemmas according to COCA (Davies, 2008). Two of Malone's three texts already met this criterion without alteration; the third text was modified by deleting rare words or replacing them with synonyms, without altering the meaning of the text, until the 97% threshold had been reached. This was done to ensure that participants could understand the texts easily. Milton and Alexiou (2009) estimated that typical intermediate and upper intermediate students have a vocabulary of between 3,000 and 4,000 lemmas, so even the least proficient participants were expected to understand most if not all of the non-target words. This was important because the evidence suggests that readers and listeners must know at least 95% (Van Zealand & Schmitt, 2013) or 96% (Waring & Takaki, 2003) of the words in a text to understand enough of the context to guess the meaning of new words. Ensuring that 97% of running words were within the 3,000 most frequent lemmas provided some leeway in case some of the non-target words were unexpectedly unknown to some participants. Malone (2018) and Chen (2021) both confirmed that the texts were comprehensible to their groups of intermediate students, and the three participants in the pilot for the present study also confirmed that they could understand the texts without difficulty.

3.3.2. Videos

Two videos were made of each text (Appendix B): one synchronous and one asynchronous. These videos were designed to be as similar as possible to avoid introducing extraneous variables. To achieve this, each text was split into sections two sentences in length. A slideshow was made by pasting each section onto a separate PowerPoint slide, written in black size 55 font on a white background. Audio files were recorded of the researcher reading each section aloud. These materials were then used to construct both videos. In the asynchronous video, the text of the first section was shown in silence and then replaced with a blank slide while the audio track of that

[1] Malone used two versions of each text: one version presented each target word four times; the other version presented each target word twice. The present study used the versions with two repetitions of each target word to ensure that the target words constituted less than 3% of the total running words.

section was played; the text of the second section was then displayed, again followed by its audio; this process continued until the end of the text. In the synchronous video, the first section of text was shown while the audio played simultaneously; this was then replaced by the second section with its accompanying audio, and this procedure was followed throughout the text. This process was inspired by Malone's (2018) study, though his participants were shown only a synchronous video, with the text split into longer sections of text.

The decision was taken to make each section of the text two sentences long. This was judged to be short enough for the asynchronous group to remember the written text while listening to the audio, which is crucial if learners are to use the orthographic text to help decode the subsequent audio (section 2.5.3), but long enough for the synchronous group to read slightly ahead of or behind the audio, which appears to be a core feature of reading-while-listening in ecologically valid contexts (Conklin et al., 2020). Since the present study was believed to be the first to investigate asynchronous reading-and-listening, no prior research could be drawn upon to evaluate the decision to use two sentence-long sections. However, the three pilot participants all reported that the format of reading and listening to short sections helped them understand the story, in both the synchronous and asynchronous conditions, which suggests that two sentence sections were suitable.

In both synchronous and asynchronous videos, the text was displayed for exactly the same length of time as the corresponding audio track took to play. The asynchronous videos were therefore twice as long as the synchronous videos, because the text and audio ran consecutively rather than concurrently.

All audio tracks included one second of silence at the start, before the text was read out, and one second of silence at the end. This was designed to give the synchronous group the opportunity to read slightly ahead of or behind the audio if desired, which Conklin et al. (2020) suggest is a key feature of synchronous reading-while-listening, without giving them enough time to read the whole text in silence, as this would have negated the synchronous condition. The asynchronous videos used the same audio track, including the silence at the start and end, to avoid introducing any unnecessary variation between the comparison groups.

The audio was recorded at a rate of approximately 100 words per minute. This was within the range of previous studies: slower than Malone's (2018) and Chen's (2021) 130 to 140 words per minute, but slightly faster than Teng's (2018) 90 words per minute. The present study adopted a slower pace than Malone (2018) and Chen (2021) to ensure that even the lowest-proficiency participants could follow the story without difficulty.

The videos were tested on a range of devices to ensure that the audio and text would be clear for participants, irrespective of which device they were using. Pilot testing confirmed that the videos were easily comprehensible, with clear audio and suitably sized text, on laptops (using both windows and apple operating systems) and smartphones (using both android and apple operating systems).

The synchronous and asynchronous participants both watched their videos once. A plausible case could have been made for the synchronous group to watch each video either once or twice. If the synchronous group had watched their videos twice while the asynchronous group watched theirs once, then each group would have had the same time-on-task, because the asynchronous video was twice as long as the synchronous one. However, requiring both the synchronous and asynchronous groups to watch their videos once ensured that each group had the same number of exposures to each word: one exposure from the written text and one exposure from the audio. Ideally, the study would have contained two synchronous groups, one adopting each option, but there were insufficient resources to do so in this study. It was therefore decided that the synchronous group should watch the video once to ensure that the number of exposures were identical in both groups.

A potential criticism should be acknowledged. The process of reading-while-listening to texts by watching videos of PowerPoint slides was very controlled and arguably lacked ecological validity: learners usually have a complete copy of the written text when reading-while-listening (Brown et al., 2008; Webb & Chang, 2012; Teng, 2018), rather than reading two sentences at a time from a video. However, this procedure was necessary to investigate the research questions properly. It was essential to split the asynchronous group's text and audio into small sections to ensure that the

participants really did read-and-listen asynchronously. To avoid extraneous variables, the text used by the synchronous group also had to be divided into sections of the same size: otherwise, any variation in outcomes might be due to the length of text presented at any one time, rather than by differences between the synchronous and asynchronous conditions. To accustom participants to the unusual experience of reading a text from a video, a short example video was provided to both groups to familiarize them with the style of their respective videos.

3.3.3. Tests

Two types of vocabulary test were used (appendices D, E, F and G), based on Malone's (2018) tests, with some modifications (sections 3.3.3.2 and 3.3.3.3). These tests were judged to meet the needs of the study and also benefitted from having been used successfully alongside the texts in two previous studies (Malone, 2018; Chen, 2021).

The tests measured receptive knowledge of orthographic form, and receptive knowledge of form-meaning links. These dimensions were targeted because they are among the first to be acquired by learners (Nation, 2022); after just two exposures to each target word in the texts, participants were expected to be in the early stages of learning the words (section 2.4.7 of the literature review). No other vocabulary tests were given, to preserve the validity of the delayed post-test as much as possible by reducing the risk of participants recognising vocabulary from the immediate post-tests rather than from the texts. This risk was unavoidable even with the two post-tests used and would have been unacceptably expanded by adding additional tests.

No vocabulary pre-test was administered because doing so could have prompted participants to learn vocabulary from the text intentionally rather than incidentally; moreover, participants could have learnt some of the target vocabulary from the pre-test as well as, or instead of, from the texts (section 2.4.2). The risk that some participants may have already known some of the target vocabulary was judged to be acceptable, as explained in section 3.3.1.

3.3.3.1 Form-meaning test

Receptive knowledge of form-meaning links was assessed by multiple choice questions; Malone's (2018) original test was judged to be suitable and was used without modification. For each target word, participants had to choose the correct definition from three options: one correct definition and two distractors. The use of three rather than four options was somewhat unusual, but it was judged to be better to maintain the existing format of the test since it had been previously used successfully. Multiple choice was a suitable format because it is easier than asking students to define words themselves (Brown et al., 2008) and is therefore better able to detect small amounts of knowledge. The distractor definitions were taken from different lexical sets to enhance the test's sensitivity to small amounts of knowledge (Brown et al., 2008). The test was scored by awarding one point for each correct answer.

Exactly the same test was used for the immediate and delayed post-tests. This was necessary to maintain the integrity of the delayed post-test: if the immediate and delayed tests had used the same correct definitions but different distractors, participants could have answered the delayed post-test by recalling which definitions they remembered from the immediate test, irrespective of whether they remembered the words from the stories.

3.3.3.2 Form recognition test

In the form recognition test, participants were given a list of the 24 target words with 24 distractors, and asked to select any words they had read or heard in the stories. The test was marked by awarding one point for every correctly selected target word and subtracting one mark for every distractor which was selected. This prevented participants from gaining full marks by guessing or asserting that they knew every word.

A different set of distractors were used in the immediate and delayed post-tests. Otherwise, participants completing the delayed post-test may have recognised distractors from the immediate post-test, and struggled to remember whether they had encountered a word in the stories or in the immediate post-test. All distractors were pseudo-words. This was a modification of Malone's original test, which used real but rare words as distractors. Pseudo-words were used to minimise the potential for participants' confusion in the delayed post-test, by ensuring that they could not

have seen any of the distractors anywhere before. Since all target words had two syllables, the pseudo-word distractors selected also had two syllables. The pseudo-words were taken from Darcy & Fontaine's (2019) list of two-syllable pseudo-words which follow the phonological rules of English.

3.3.3.3. Comprehension tests

Participants were given reading comprehension questions (Appendix C) to focus them on the meaning of the texts (Malone, 2018) and thereby ensure that the learning condition was incidental (section 2.1). The comprehension questions were used only to focus participants on the meaning of the texts, and to check that they had understood at least some key features of the stories. Malone used eight comprehension questions per story. To save time, only four questions per story were used in the present study. The comprehension questions developed by Malone were judged to be suitable for these purposes, so four of the original eight questions were used without modification for each story.

3.3.4. The final product

The videos, comprehension questions, immediate vocabulary tests, and instructions were brought together into one webpage which guided participants through every stage of the process. There were two versions of this webpage: one with synchronous videos and one with asynchronous videos. Apart from containing different videos, these two versions were identical in every respect. A third webpage was created to guide participants through the delayed vocabulary tests.

The webpages were made using google forms. Each webpage had clear instructions introducing each activity. Participants first entered their names and dates of birth and then followed instructions to watch the videos and answer the test questions. More details about the procedure can be found in section 3.4. The webpages were tested on a range of devices, and were found to be easy to use on laptops (running both windows and apple operating systems) and phones (using both android and apple operating systems).

3.4. Procedure

Participants completed a short background information survey prior to joining the first session (appendix K), covering their level of education, gender, date of birth, L1, and the number of years they had been studying English. Participants were given only general information about the research aims in advance of the study (appendix I): they were told that the aim was to investigate how reading and listening helped people like them improve their English. Vocabulary learning was not mentioned, to maintain the incidental learning condition (section 2.1).

Following this, they took part in two sessions, both conducted online using videoconferencing software. In the first session, participants watched the videos and then took the immediate vocabulary post-tests; in the second session, participants took the delayed vocabulary tests. Participants thus attended only one reading-and-listening session. It would have been desirable to have a longer intervention in which participants attended multiple sessions and watched more videos in total. However, this was not feasible given the time and resources available. It was judged that one reading-and-listening session would be sufficient because three of the six previous studies into incidental vocabulary acquisition from reading-while-listening successfully used only one treatment session, and all three detected vocabulary learning (Teng, 2018; Malone, 2018; Chen, 2021).

All sessions were led by the researcher. The number of participants in each session varied from one to six, but this variation was not anticipated to affect participants' performance because the actual treatments were carried out individually, as explained below.

3.4.1. Session One

The first session took around 50 minutes for the synchronous group and 75 minutes for the asynchronous group. After a short plenary introduction, each participant was sent a link to a webpage (appendix H), which guided them through the experimental activities in the following order:

Participants first watched a short example video followed by one comprehension question. For the synchronous group, the video presented a short text with synchronous text and audio; for the

asynchronous group, the video was asynchronous. This example was intended to identify any initial technical issues and accustom participants to the format of the videos.

Participants were then asked to watch the first video. They were directed to focus on understanding the video and finding the answers to the associated comprehension questions. The relevant comprehension questions were displayed at the start of each video to focus participants on the meaning of the texts and thus maintain incidental learning conditions. Participants then watched the video of the story, without pausing or rewinding it unless they encountered technical problems. After watching the video, they attempted to answer the comprehension questions. This process was repeated for the other three texts.

Surprise vocabulary post-tests were then administered. To maintain incidental learning conditions, participants had not been pre-warned about these tests. The form recognition test was given first, followed by the form-meaning test. This was important because if participants had taken the form-meaning test first, this would have given them another opportunity to notice the form of the target words, which could then have boosted their scores in the form-recognition test. Finally, participants submitted the data from the webpage and had a brief verbal debrief with the researcher.

Throughout this process, participants were aware that if they had any technical problems, they were welcome to try to fix them themselves initially (for example by replaying the video if there had been a problem with the audio) and they could also call the researcher for help, who was on hand at all times.

It was important to ensure that participants took part in the activity exactly as intended, without taking notes, watching the video multiple times, or looking up the words for the form-meaning test. Several steps were taken to achieve this. First, participants were explicitly instructed in the initial plenary introduction to watch each video once only, without pausing or rewinding, to take no notes, and to answer all questions without looking anything up, using only their memory of the videos. Second, after completing the activities, participants were asked to confirm whether they had looked up any of the words for the form-meaning test. Third, the researcher monitored the time that each participant took to complete the activities to confirm that they had not had time to watch the videos more than once.

3.4.2. Session Two

Session two took place seven days after session one, and lasted for approximately fifteen minutes. After a brief plenary introduction, participants received links to a webpage on which they took the delayed post-tests. Again, the form recognition test was taken first, followed by the form-meaning test, and the researcher was available to assist with any technical problems.

There was a risk that participants might look up or study some of the target words before the delayed post-test. To mitigate this, participants were asked at the end of the first session not to look up or study any of the words until after the delayed post-test. At the beginning of session two, participants were asked to confirm whether they had studied any of the target words; all stated that they had not.

3.5. Pilot

3.5.1. Procedure

A pilot was conducted to check that the materials and procedure were fit for purpose and identify any potential technical problems. This was a technical run-through rather than a full pilot: only three participants took part and no inferential statistical analysis was carried out of the results. Three participants were recruited to take part: one man and two women, all aged 19 with a B1 level of English. All three had been recruited through teacher 1 and were students at the university in Iraq. Ideally, it would have been better to involve participants from a wider range of levels and centres in the pilot; however, this was not possible due to time constraints: at the time that the pilot was conducted, the only participants who had been recruited were B1 level students at the Iraqi university. However, this was not judged to be a serious limitation because the greatest concern with running the pilot was to ensure that all participants understood the videos and instructions clearly; B1 level students were judged to be at most risk of not understanding because their English level was lowest, so they were judged to be the most useful group to take part in the pilot.

Initially, the pilot was conducted with two participants: both followed the entire procedure outlined in section 3.4, one using the synchronous materials and the other using the asynchronous materials. This procedure revealed a problem with the instructions for one of the tests, as described below. In

response, the instructions were rewritten and the updated materials were trialled on one further participant (who used the synchronous materials); this confirmed that the modified instructions were clearly comprehensible.

Data were collected from the pilot in three ways. First, the researcher took notes during the session, focusing on technical issues, timing, and any other emergent issues. Second, the participants' test scores were evaluated to confirm that they were able to engage with the activities appropriately. Third, after completing the procedure, participants were engaged in a group discussion to evaluate the materials. The discussion was guided by the questions outlined in appendix L, but participants were encouraged to share their own thoughts freely on any topic. This semi-structured approach was designed to enable unexpected problems or issues to emerge. The data were triangulated to evaluate the materials and procedure.

3.5.2. Results of the pilot

The pilot confirmed that the materials and procedure were suitable for purpose after the two minor alterations described below. Participants reported that the videos were easy to understand, and all three of them scored 100% on the comprehension questions. Participants reported that they had enough time to read the text properly, and reported that both synchronous and asynchronous formats were helpful and facilitated their understanding of the story. Technically, everything worked well, with the exception of a minor malfunction in one video, detailed below: participants reported that the videos, webpage and tests were easy to use. Most instructions were clear: participants found the instructions for the videos, the comprehension questions, and the form-meaning test to be clear, and they completed these tasks as intended. The participants also confirmed that almost all of the target vocabulary was new to them: two participants had not encountered any of the target words before, while one participant had previously encountered just one of the 24 target words (shyster).

Two problems were identified, and modifications were made to address them. First, the two original participants both misunderstood the instructions for the form recognition tests. These instructions were rewritten to make them clearer, and the third pilot participant understood the modified

instructions without any difficulty. Second, a small malfunction was identified in one of the asynchronous videos; in response, the video was remade to fix the issue.

3.6. Ethical considerations

Informed consent was obtained from all participants. Before completing the consent forms (appendix J), participants were given information about the study (appendix I) and had the opportunity to ask the researcher additional questions. As discussed in section 3.4, participants were given only general information about the aims of the study, and vocabulary learning was not mentioned, in order to maintain an incidental learning condition. However, all the information given to participants was accurate: no deception was used. After completing the second session, participants were given a more detailed verbal explanation of the research objectives.

Data were collected using google forms, stored temporarily on a password-protected google account, then transferred to a password protected Oxford University OneDrive account. Once all data had been collected, they were anonymized and the originals, which included participants' names, was destroyed. The data were accessible only by the researcher and their supervisor at the University of Oxford. Participants were aware that they could withdraw from the study at any time before taking part, and that they could withdraw their data until it had been anonymised.

Two incentives were offered to encourage participants to join the study. First, they had the opportunity to join a free online English lesson delivered by the researcher. Second, they had the opportunity to be entered into a prize draw, in which four participants were randomly chosen to win $50 each in gift tokens. Participants were free to opt out of both incentives if they so desired.

4. Results

4.1. Data analysis

All statistical analyses were carried out using SPSS version 29.0.1.0(171). The data was analysed to check whether it met the assumptions required to conduct t-tests, and non-parametric alternatives were used when these assumptions were not met.

4.2. Participants and exclusions.

Twenty-eight participants were recruited and took part in session 1. The results of one participant were excluded from all analysis because he followed the procedure incorrectly, by watching only one of the three videos and then skipping ahead to answer the immediate vocabulary tests. His results were therefore excluded from all analysis, leaving 27 participants, who are described in section 3.2.2.

Of these 27 participants, the results of one were partially excluded because she stated in the debrief session following the immediate vocabulary tests that she had looked up the answers to the form-meaning test online. Her results for the form-meaning test were therefore excluded, and she was not invited to take the delayed post-tests because her performance could have been affected by the fact that she had looked up the words. However, her results for the comprehension questions and form recognition test were retained, for three reasons. First, she stated that she had carried out these tests as instructed. Second, the only practical way to cheat on these tests would be to re-watch the videos, and she had not had time to do this. Third, participants completed the comprehension and form-recognition tests before taking the form-meaning test, so the fact that she looked up the answers for the form-meaning test could not have impacted her answers on these earlier tests.

The remaining 26 participants all confirmed in the debrief that they had followed the instructions as intended. Their results were therefore retained, and all were invited to attend the delayed post-tests. Six of these then dropped out: five dropped out from the synchronous group, while one dropped out from the asynchronous group. As a result, 20 participants took the delayed vocabulary tests. All 20 of these participants confirmed that they had not looked up any of the words in the week between the immediate and delayed tests, and after taking the delayed tests, they all

confirmed that they had followed the procedures correctly. Therefore, the results of all 20 were retained.

In summary, results were retained from 27 participants for the comprehension questions and the immediate form recognition test. For the immediate form-meaning test, results were retained from 26 participants. For the delayed form-meaning and form-recognition tests, results were retained from 20 participants.

Tables 2, 3 and 4 compare the participants in the comparison groups at each of these three points, presenting data on the number of participants and their proficiency levels. The comparison groups were very similar to each other at the immediate tests, because stratified random allocation had been used. For the delayed vocabulary tests, the groups differed more sharply in size because more participants dropped out of the synchronous group; however, both groups retained a similar range of proficiency levels.

Table 2: participants whose data was retained for the comprehension and immediate form-recognition tests

	N	**Proficiency levels**
Synchronous group	13	B1 (n=3); B2 (n=8); C1 (n=2)
Asynchronous group	14	B1 (n=3); B2 (n=8); C1 (n=2); B1/B2 (n=1)

*The teacher believed this student to be either B1 or B2 but could not be more specific

Table 3: participants whose data was retained for the immediate form-meaning test

	N	**Proficiency**
Synchronous group	12	B1 (n=2); B2 (n=8); C1 (n=2)
Asynchronous group	14	B1 (n=3); B2 (n=8); C1 (n=3); B1/B2* (n=1)

*The teacher believed this student to be either B1 or B2 but could not be more specific

Table 4: participants whose data was retained for the delayed vocabulary tests

	N	Proficiency
Synchronous group	7	B1 (n=2); B2 (n=4); C1 (n=1)
Asynchronous group	13	B1 (n=3); B2 (n=8); C1 (n=2).

4.3. Comprehension tests

For the comprehension questions, descriptive statistics were calculated, but no inferential statistical analyses were conducted. This was because the comprehension tests did not relate directly to either research question: their purpose was simply to focus the participants on the meaning of the stories, and to check that the stories had been adequately understood.

The average score of the 27 participants on the comprehension tests was 11.30 out of a possible 12 (94%). The lowest score was 8 out of 12 (67%); this score was obtained by one participant. The other 26 participants all scored at least 10 out of 12 (83%). This suggests that the participants had understood the stories.

4.4. RQ1

To address RQ1, the performance of the synchronous vs. asynchronous groups on the immediate vocabulary tests were compared using independent-sample t-tests. Two t-tests were run: one on the form-recognition test scores, and one on the form-meaning test scores. Since H0 and H1 were one-tailed (section 2.6.2), one-tailed t-tests were conducted.

4.4.1. Immediate form-recognition test

Descriptive statistics of scores on the immediate form-recognition test are presented in table 5 below. The maximum score available was 24.

Table 5: Form-recognition: Descriptive statistics

Group	N	Mean	Standard Deviation	95% confidence interval for mean
Synchronous	13	8.92	6.61	4.93 to 12.92
Asynchronous	14	11.86	6.24	8.26 to 15.46

Analysis revealed that the assumptions required to conduct independent t-tests were all met, and a one-tailed independent-sample t-test was therefore conducted. The asynchronous group had a larger mean with an effect size of d=.46, which is small according to Plonsky and Oswald's (2014) guidelines. However, no statistically significant difference between the comparison groups was detected (p=.123). The results are presented in table 6.

Table 6: Form recognition: t-test

T	DF	Mean difference	95% confidence of group difference	P	D
1.19	25	2.93	-2.16 to 8.03	.123	.46

4.4.2. Immediate form-meaning test

Descriptive statistics of scores on the immediate form-meaning test are presented in table 7 below. The maximum score available was 24.

Table 7: Form-meaning: Descriptive statistics

Group	N	Mean	Standard Deviation	95% confidence interval for mean
Synchronous	12	15.75	4.00	13.21 to 18.29
Asynchronous	14	18.29	3.07	16.51 to 20.06

Again, analysis revealed that the assumptions required to conduct independent t-tests were all met and a one-tailed independent-sample t-test was therefore conducted. The asynchronous group had a larger mean with an effect size of d=.72, which is regarded as medium by Plonsky and Oswald (2014). The difference between the groups was statistically significant (p=.04). The results are presented in table 8.

Table 8: Form-meaning: t-tests

T	DF	Mean difference	95% confidence of group difference	P	D
1.83	24	2.54	-.33 to 5.40	.04	.72

4.5. RQ2: Delayed vocabulary tests

Several tests were conducted to address RQ2 and explore the retention of vocabulary gains over time. First, the performance of the synchronous vs. asynchronous groups on the delayed vocabulary tests were compared using independent-sample t-tests, and non-parametric alternatives when necessary. Again, two tests were run: one for form-recognition scores, and one for form-meaning scores. Following the formulation of H0 and H1 as one-tailed hypotheses, one-tailed t-tests were conducted.

4.5.1. Delayed form-recognition test

Descriptive statistics of scores on the delayed form-recognition test are presented in table 9 below. The maximum score available was 24.

Table 9: Form-recognition: descriptive statistics

Group	N	Mean	Standard Deviation	95% confidence interval for mean

Synchronous	7	10.71	6.80	4.43 to 17.00
Asynchronous	13	8.92	6.99	4.70 to 13.15

The assumptions required to conduct independent t-tests were not met; a Mann-Whitney U-test was therefore conducted. A visual inspection determined that the distribution of scores for the synchronous and asynchronous groups was not similar. There was no statistically significant difference in scores between the two groups, U=41.50, z=-.32, p=.757.

4.5.2. Delayed form-meaning test

Descriptive statistics of these participants' scores on the form-meaning test are presented in table 10 below. The maximum score available was 24.

Table 10: Delayed form-meaning: descriptive statistics

Group	N	Mean	Standard Deviation	95% confidence interval for mean
Synchronous	7	16.86	3.93	13.22 to 20.50
Asynchronous	13	16.00	4.86	13.06 to 18.94

The assumptions required to conduct independent t-tests were met, and a one-tailed independent-sample t-test was therefore conducted. The synchronous group had a larger mean with an effect size of d=.19, which is regarded as small by Plonsky and Oswald (2014). The difference between the groups was not statistically significant (p=.35). The results are presented in table 11.

Table 11: Delayed form-meaning test: t-test

T	DF	Mean difference	95% confidence of group difference	P	D
4.00	18	0.86	-3.65 to 5.36	.347	.19

4.5.3. Within-group vocabulary retention

Next, the extent to which each group retained vocabulary knowledge between the immediate and delayed vocabulary tests was calculated. Paired t-tests, or non-parametric alternatives, were conducted to compare the immediate and delayed vocabulary test scores within each comparison group. These tests were carried out to establish whether either of the groups displayed a statistically significant change in scores between the immediate and delayed vocabulary tests. All t-tests were two-tailed because no one-tailed hypotheses had been formulated.

Only the data of the 20 participants who took both the immediate and delayed vocabulary tests was included in this analysis: participants who took the immediate tests but not the delayed tests was excluded. Descriptive statistics of these test scores are presented in table 12.

Table 12: Immediate and delayed test scores: descriptive statistics

Group	**Test**	**N**	**Immediate test scores**			**Delayed test scores**		
			Mean	**SD**	**95% confidence of mean**	**Mean**	**SD**	**95% confidence of mean**
Sync.	Form-recognition	7	11.14	6.18	5.43 to 16.85	10.71	6.80	4.43 to 17.00
Sync.	Form-meaning	7	16.71	3.50	13.48 to 19.95	16.86	3.93	13.22 to 20.50
Async.	Form-recognition	13	11.31	6.13	7.60 to 15.01	8.92	6.99	4.70 to 13.15
Async.	Form-meaning	13	18.00	3.00	16.19 to 19.81	16.00	4.86	13.06 to 18.94

For the asynchronous group, the scores on both the form-recognition and form-meaning tests met the assumptions required for a paired t-test. For the synchronous group, the form-recognition test scores met the necessary assumptions for a paired t-test but the form-meaning test scores did not. Two-tailed, paired t-tests were therefore conducted on the three sets of scores for which the necessary assumptions had been met (table 13). The asynchronous group displayed a statistically significant decrease in scores on the form-meaning test between the immediate and delayed tests (p=.029). The effect size (d=.67) was medium according to Plonsky and Oswald's (2014) criteria. No other results were statistically significant.

Table 13: Immediate vs. delayed vocabulary test scores: paired t-tests

Group	Test	T	DF	Mean difference	95% confidence of group difference	P	D
Sync.	Form-recognition	.32	6	.43	-2.90 to 3.76	.763	.12
Async.	Form-recognition	2.13	12	2.38	-.05 to 4.82	.054	.59
Async.	Form-meaning	2.47	12	2.00	.24 to 3.76	.029	.67

A Wilcoxon signed-rank test was conducted on the synchronous group's form-meaning test scores. A visual inspection of a histogram confirmed that the distribution was symmetrical and that the Wilcoxon signed-rank test was therefore appropriate. Of the seven participants, two saw an increase in their scores from the immediate to the delayed test; two saw a decrease in their scores; for three, their scores remained unchanged. The synchronous and asynchronous groups had the same median: 19. There was no significant difference between the medians of the synchronous group (median = 19) and the asynchronous group (median=19), z=.18, p=.854.

4.5.4. Between-group vocabulary retention

Finally, tests were conducted to compare the two comparison groups on the extent to which they retained or forgot vocabulary between the immediate and delayed tests. For each participant, the difference between the immediate and delayed vocabulary test scores was calculated. Independent t-tests or non-parametric equivalents were then conducted to compare the amount of vocabulary lost by the synchronous vs. asynchronous groups between the two tests. Again, since no specific hypotheses had been formulated, two-tailed t-tests were used. Only participants who had completed both the immediate and delayed tests were included in this analysis.

First, the difference between each participant's scores on the immediate and delayed tests was calculated. For each participant, subtracting their delayed score from their immediate score revealed the number of words they had forgotten between the immediate and delayed tests. Table 14 presents descriptive statistics of these differences. In both cases, the asynchronous group had a larger mean difference, meaning that more words were forgotten by this group.

Table 14: Between-group vocabulary retention (immediate score – delayed score): descriptive statistics

Test	Group	N	Mean difference (immediate score – delayed score)	Standard Deviation	95% confidence interval for mean
Form-recognition	Sync.	7	.43	3.60	-2.90 to 3.76
	Async.	13	2.38	4.03	-.05 to 4.82
Form-meaning	Sync.	7	-.14	2.12	-2.10 to 1.81
	Async.	13	2.00	2.92	.24 to 3.76

The differences on the form-recognition test met the assumptions required to perform independent t-tests; a two-tailed independent t-test was therefore conducted (table 15). The synchronous group had a larger mean with an effect size of d=.50, which is regarded as small by Plonsky and Oswald (2014). However, the difference between the groups was not statistically significant (p=.30).

Table 15: Between group vocabulary retention (immediate score-delayed score) on the form-recognition test: t-test

T	DF	Mean difference	95% confidence of group difference	P	D
1.07	18	1.96	-1.88 to 5.79	.298	.50

The differences on the form-meaning test did not meet the assumptions required for independent t-tests, so a Mann-Whitney U-test was therefore conducted. A visual inspection determined that the distribution of scores for the synchronous and asynchronous groups was not similar. There was no statistically significant difference in scores between the two groups, U=65.00, z=-.156, p=.135.

4.6. Summary

In summary, no statistically significant differences between groups were detected on the immediate form recognition test, suggesting that student's ability to recognise a word were not influenced by the type of instruction. There was a statistically significant difference between groups on the immediate form-meaning test, in favour of the asynchronous group (p=.04). This suggests that asynchronous reading-and-listening was associated with better retention of the meaning of the target words.

No statistically significant differences between groups were detected for either the delayed form recognition test or the delayed form-meaning test. This suggests that the type of instruction did not influence participant's long-term retention of either dimension of vocabulary knowledge.

Within the synchronous group, there were no statistically significant differences between the immediate and delayed scores on either test, suggesting that little vocabulary was forgotten during the intervening week. Within the asynchronous group, there were no statistically significant differences between the immediate and delayed scores on the form-recognition test, suggesting that these participants retained the ability to recognise the words they had learnt. However, participants in the asynchronous group gained statistically significantly lower scores on the delayed form-meaning test than on the immediate form-meaning test; this suggests that these participants had forgotten some of the words during the intervening week.

Finally, there were no statistically significant differences between the groups in the rate at which vocabulary was forgotten between the immediate and delayed tests. This suggests that the type of instruction did not influence the speed at which participants forgot vocabulary.

5. Discussion

5.1. RQ1: Immediate effects

5.1.1. Between-group differences on the immediate vocabulary test scores

RQ1 focused on the relative effects of synchronous reading-while-listening vs. asynchronous reading-and-listening on incidental vocabulary acquisition, as measured by immediate vocabulary tests. Analysis of test scores revealed that the asynchronous group performed statistically significantly better than the synchronous group on the immediate form-meaning test ($p=.04$, $d=.72$), though there was no statistically significant difference between the groups on the immediate form-recognition test. H_1 was therefore confirmed for the form-meaning test, while H_0 was retained for the form-recognition test. This suggests that the asynchronous approach is more effective than the synchronous approach in facilitating at least some forms of vocabulary knowledge under incidental learning conditions. This result is noteworthy because the present study is, as far as I am aware, the first to investigate the impact of asynchronous reading-and-listening on incidental vocabulary acquisition; this is therefore the first piece of evidence on the effects of this approach.

5.1.2. Why might asynchronous reading-and-listening perform better than synchronous reading-while-listening?

As suggested in the literature review (section 2.5.3), an explanation for why the asynchronous approach was more effective than the synchronous approach may lie in their relative effects on cognitive load. Cognitive load theorists (Diao and Sweller, 2007; Moussa-Inaty et al., 2012) have argued that synchronous reading-while-listening requires learners to pay simultaneous attention to both phonological and orthographic input, and use some of their finite working memory capacity to integrate the two. This creates extraneous cognitive load, reducing the amount of working memory available to carry out other tasks such as understanding the text and processing the meaning of new vocabulary. In turn, this is likely to limit the amount of vocabulary learnt, both directly, by reducing the capacity of working memory to integrate a new word with existing schemata (Moussa Inaty et al., 2012), and indirectly, by limiting learners' understanding of the text (Conklin et al., 2020), which then curtails their ability to learn words from context (Nation, 2022).

Moreover, Conklin et al.'s (2020) study suggests that people who are reading-while-listening typically read ahead of the audio (Conklin et al., 2020). The resultant misalignment between aural and textual input may bring both advantages and disadvantages. The process of reading ahead of the audio could benefit learners by helping them segment and understand the subsequent aural input, enabling them to link phonological and orthographic forms more effectively (Conklin et al., 2020). In fact, this mechanism may explain why synchronous reading-while-listening is an effective approach (section 2.5.1). However, the need to process the streams of aural and textual input simultaneously rather than sequentially may generate extraneous cognitive load, because simultaneously focussing on different words in the audio and orthographic input is likely to place a heavy demand on working memory. Asynchronous reading-and-listening reduces this extraneous cognitive load because learners only have to focus on one stream of input at a time: their attention is not split between aural and orthographic input. As a result, more working memory resources are available to focus on other processes which help transfer the new vocabulary to long term memory, such as inferring meaning from context. Asynchronous reading-and-listening may therefore bring the same key benefits as synchronous reading-while-listening, by allowing learners to read ahead of and segment the audio input, while avoiding its main disadvantage by imposing a lower extraneous cognitive load.

However, it is important to recognise that other explanations for the overperformance of the asynchronous approach are also possible. For example, the relative advantage of the asynchronous group could also be explained by the fact its participants had more time-on-task (Nation, 2007). Further research is needed to investigate which mechanism best explains these results (section 6.3). On a similar note, the relative advantage of the asynchronous approach cannot be assumed to apply to learners of all levels in all contexts. The present study involved learners at an intermediate level or higher; more research is needed to investigate the relative effects of asynchronous vs. synchronous approaches on learners at lower proficiency levels (section 6.3).

It is interesting that no statistically significant difference between the groups was detected on the immediate form-recognition test. This could simply be because the present study lacked the statistical power required to detect such a difference. Alternatively, it is possible that the relative advantages of the asynchronous approach apply only to some dimensions of vocabulary knowledge, including linking form and meaning, and not to other dimensions, including form recognition. This could also be explained by cognitive load theory. A key tenant of cognitive load theory is that

information must be processed in working memory in order to be learnt (Moussa-Inaty et al., 2012). I have argued above that the synchronous approach requires learners to focus more of their cognitive resources on processing and integrating written and aural word forms, which leaves fewer resources available to process meaning. The synchronous approach thus affords learners the opportunity to process word form at the expense of processing word meaning; it should therefore facilitate knowledge of word form relatively more effectively than it facilitates knowledge of meaning.

5.1.3. Other observations about the immediate vocabulary test scores

Moving away from direct comparisons between the groups, it is also noteworthy that both groups gained much higher scores on both tests than would be expected by chance. This suggests that both approaches were effective facilitators of incidental vocabulary acquisition, even though participants had only two exposures to each target word. A participant with no knowledge of any of the target words would have been expected to score an average of zero out of 24 on the form-recognition test, and eight out of 24 on the form-meaning test, if they had simply guessed the answers to each question. In contrast, the synchronous group scored a mean of 8.92 on form-recognition and 15.75 on form-meaning. This result corroborates the evidence accumulated by previous studies into synchronous reading-while-listening (Chen, 2021; Teng, 2018; Malone, 2018; Webb & Chang, 2012) that it is an effective technique that can facilitate useful gains in vocabulary under incidental learning conditions (section 2.4.7). The asynchronous group also performed better than would be expected by chance, scoring a mean of 11.86 on the form-recognition test and 18.29 on the form-meaning test. This supports the conclusion that asynchronous reading-and-listening is also an effective approach for incidental vocabulary learning.

5.2. RQ2: delayed effects

RQ2 investigated the extent to which any gains in vocabulary knowledge following synchronous reading-while-listening and asynchronous reading-and-listening were retained over time. This question can be divided into several parts. First, did the asynchronous group outperform the synchronous group on the delayed vocabulary tests? Second, to what extent did each group retain

their vocabulary gains between the immediate and delayed tests? Third, were there any differences between the groups in the extent to which they retained vocabulary between the immediate and delayed tests? These questions are explored in more detail below.

5.2.1. Between-group differences in the delayed vocabulary test scores

Analysis of the delayed vocabulary test results identified no significant differences between the scores of the synchronous vs. asynchronous groups. H_0 was therefore retained: although the asynchronous approach appeared to prompt greater immediate gains on some dimensions of vocabulary knowledge (section 5.1.1), there is no evidence that this advantage persisted one week after the treatment.

It is important not to generalise overconfidently from these results, because they are limited by the high drop-out rate among participants in the synchronous group. Of the thirteen participants who took the synchronous treatment, only seven returned to complete the delayed vocabulary test; in contrast, 13 of the 14 asynchronous participants took the delayed vocabulary test. This high drop-out rate had the effect of reducing the sample size for the synchronous group. Moreover, the participants who dropped out may have differed systematically from those who took the delayed test and this could potentially have influenced the results. For example, participants who found the experience useful and felt that they had learnt something from it may have been more likely to return. Therefore, the delayed test results for the synchronous group may not give an accurate impression of the impact of this approach on the average participant. Nevertheless, it could be the case that the asynchronous approach really does outperform the synchronous approach in the short term but not the long term. It is unclear why this may be, but perhaps long-term vocabulary retention depends less on the medium in which the words were initially encountered, and more on the extent to which participants thought about the story and the words therein during the week after the treatment, which may not be as influenced by the medium in which the story was originally presented.

5.2.2. Vocabulary retention in the synchronous group

For the synchronous group, mean scores on the delayed tests were very similar to those on the immediate tests, with no statistically significant differences between immediate and delayed test scores for either the form-recognition or the form-meaning tests. This indicates that very little vocabulary was forgotten during the intervening week. These results must again be interpreted with caution because of the high drop-out rate in the synchronous group (section 5.2.1), but they are noteworthy because they corroborate the findings of the two previous reading-while-listening studies which included delayed vocabulary tests. Brown et al. (2008) detected no statistically significant loss of vocabulary knowledge on multiple choice tests of meaning. Vu and Peters (2020) did not calculate the statistical significance of differences between immediate and delayed vocabulary tests, but their descriptive statistics indicate that scores only declined slightly between the immediate and delayed tests. However, as argued in the literature review, both studies failed to establish genuinely incidental vocabulary learning conditions: Brown et al. (2008) because participants were explicitly asked to memorise unfamiliar vocabulary (section 2.4.1), and Vu and Peters (2020) because participants had good reasons to expect vocabulary post-tests (section 2.4.5). The present study therefore makes a novel contribution by detecting a similar effect under truly incidental learning conditions. This strengthens the evidence base for the effectiveness of synchronous reading-while-listening by suggesting that the incidental vocabulary gains it facilitates can be retained over time.

Drilling down into the test results in more detail reveals that while the mean score on the form recognition test fell slightly (and non-statistically significantly) between the immediate and delayed tests, the mean score on the form-meaning test actually rose slightly from 15.75 on the immediate test to 16.86 on the delayed test. This is somewhat surprising, though previous studies into incidental vocabulary learning have also found that some learners scored better on delayed than immediate tests (Brown et al., 2008; Webb et al., 2023). This finding should not be over-interpreted as it could well have been caused by chance (Brown et al., 2008): the form-meaning test was a multiple choice test which allowed participants to guess, so some random variation in scores is to be expected. Another possible explanation, proposed by Brown et al. (2008), is that participants who gained partial knowledge of the vocabulary from the initial treatment may have been able to deepen this knowledge during the week before the delayed test by relating this vocabulary to other words in their lexicon.

5.2.3. Vocabulary retention in the asynchronous group

In the asynchronous group, mean scores declined somewhat from the immediate to the delayed vocabulary tests: they fell from 18.00 to 16.00 on the form-meaning test, and from 11.31 to 8.92 on the form-recognition test. Paired t-tests revealed that this vocabulary loss was statistically significant for the form-meaning test ($p=.029$, $d=.67$), but not for the form-recognition test. The scores on both tests remained far above what would have been expected by chance, indicating that the majority of vocabulary learnt from the texts had been retained after one week. It should be noted that the drop-out rate for the asynchronous group was much lower than for the synchronous group: 13 of the original 14 participants took the delayed test; as a result, these findings can be interpreted with more confidence.

Since the present study is the first to investigate incidental vocabulary acquisition from asynchronous reading-and-listening, there are no directly equivalent studies to compare these results to. However, the overall picture of a modest decline in test scores over time is in line with the findings of some studies into incidental vocabulary learning through other modes of input. For example, Chen and Truscott's (2010) study into incidental vocabulary acquisition from reading only also found that mean scores on most tests declined between the immediate and delayed vocabulary tests (section 2.3). As discussed in section 5.2.2, studies into synchronous reading-while-listening (Brown et al., 2008; Vu & Peters, 2020) similarly detected declines from immediate to delayed tests.

5.2.4. Between-Group Differences in Vocabulary Retention

Finally, the groups were compared to check if there were any statistically significant differences in the extent to which they retained vocabulary between the immediate and delayed vocabulary tests. No statistically significant differences were detected: even though the asynchronous group displayed a statistically significant decline in scores while the synchronous group did not, there were no statistically significant differences when comparing the two groups. Therefore, there is no evidence to determine whether the synchronous or asynchronous approaches led to more durable knowledge.

6. Conclusion

6.1. Research Questions re-examined

This dissertation was motivated by two research questions. These research questions are restated below, and key conclusions are set out.

RQ1: *What are the relative effects of synchronous reading-while-listening and asynchronous reading-and-listening on incidental vocabulary learning, as measured by vocabulary tests administered immediately after the treatment?*

The results suggest that asynchronous reading-and-listening facilitated more incidental vocabulary acquisition than synchronous reading-while-listening in the short term, on at least some dimensions of vocabulary knowledge. This is demonstrated by the fact that the asynchronous approach statistically significantly outperformed the synchronous approach on the immediate form-meaning test, though there were no statistically significant differences between the groups on the immediate form-recognition test. A plausible explanation for the greater effectiveness of asynchronous reading-and-listening is that it brings the same benefits as synchronous reading-while-listening, while reducing extraneous cognitive load.

RQ2: *To what extent are any gains in vocabulary knowledge following synchronous reading-while-listening and asynchronous reading-and-listening maintained over time, as measured by delayed vocabulary tests?*

There is no evidence that asynchronous reading-and-listening facilitates more incidental vocabulary acquisition than synchronous reading-while-listening in the longer term, because no statistically significant differences between groups were detected in the delayed vocabulary tests. However, within-group comparisons of immediate and delayed vocabulary test results suggest that both approaches can promote durable vocabulary knowledge. After one week, participants in the asynchronous group still remembered most of the vocabulary they had learnt, though there was a statistically significant drop in scores between the immediate and delayed form-meaning tests. Participants in the synchronous group retained the vast majority of vocabulary they had learnt, with no statistically significant declines between immediate and delayed test scores on either dimension of vocabulary knowledge. This suggests that both approaches are effective facilitators of incidental vocabulary acquisition in the longer term.

6.2. Pedagogical Implications

More research is needed to corroborate or contest the findings of this study (section 6.3). Nevertheless, some provisional pedagogical implications can be drawn.

The results suggest that language learners should be encouraged to add asynchronous reading-and-listening to their repertoire of language learning activities, because it appears to be an effective approach which can enhance the incidental acquisition of vocabulary. The asynchronous approach could be implemented in classrooms, as a teacher-led activity; alternatively, students could be encouraged to read-and-listen asynchronously at home, perhaps as part of a broader extensive reading program. It is worth noting that while the present study used specially developed videos to facilitate asynchronous reading-and-listening, these would not be required in normal learning conditions. Learners could administer the asynchronous approach using any reading materials which have an associated audio track, by simply reading a small part of the text and then playing the corresponding section of the audio. Videos were used in the present study to control the participants' behaviour and ensure that the two conditions were as similar and controlled as possible, enabling valid comparisons between the synchronous and asynchronous approaches to be made; such videos are not an essential prerequisite of asynchronous reading-and-listening in general.

It was suggested in section 5.1.2 that asynchronous reading-and-listening helps reduce extraneous cognitive load. If this is correct, the asynchronous approach may be especially useful for students who would otherwise experience unusually high levels of cognitive load. Thus, students who struggle to decode aural input, or are reading texts containing language they find challenging, may find the asynchronous approach to be particularly useful.

Learners are liable to forget some of the vocabulary they acquire through the asynchronous approach; to prevent this, the vocabulary they learn should be revisited. The process of revising vocabulary could take many forms (Nation, 2022): learners could be given activities which require them to manipulate or produce the target vocabulary, study the new vocabulary using flashcards, or follow an extensive reading program in which newly learnt words are likely to re-occur in other texts.

The present study also demonstrated that vocabulary learnt from synchronous reading-while-listening can be retained over time. This strengthens the evidence from previous studies that synchronous reading-while-listening is also a useful approach (Chen, 2021; Malone, 2018; Teng, 2018; Webb & Chang, 2012). The synchronous and asynchronous approaches both appear to be effective; the asynchronous approach promoted more learning in the short term, but no long term between-group differences were detected. Therefore, both approaches can play a role in language learning. When choosing between them, educators should draw on evidence about their relative effectiveness, alongside other factors such as learner preferences and the practicalities of course and lesson planning.

6.3. Limitations and recommendations for further research

The present study is believed to be the first to investigate incidental vocabulary acquisition through asynchronous reading-and-listening; as a result, replications are needed to assess the veracity of its conclusions. In addition, future studies should investigate whether similar effects obtain in different contexts, by recruiting participants who have different L1s and speak English at different proficiency levels.

Further research could also improve on the present study by using a larger sample size. In this dissertation, twenty-seven participants were recruited and divided into two comparison groups: 14 participants in one group and 13 in the other. These comparison groups were larger than those employed by several previous studies into reading-while-listening (Brown et al., 2008; Vu & Peters, 2020), and provided enough statistical power to detect some statistically significant effects. Nevertheless, it would have been desirable to have a larger sample size and a corresponding increase in statistical power. The original aim was to recruit approximately 50 participants, 25 per experimental group, but this proved to be impossible because of a lack of necessary resources. Future studies with more resources at their disposal should aim to include a larger number of participants.

Another limitation of the present study was the high drop-out rate among participants in the synchronous group, which reduced the confidence in the conclusions drawn from the delayed

vocabulary test scores. Future studies would benefit from a higher retention rate if this can be achieved, perhaps by offering additional incentives to participants who attend both sessions.

As noted in section 5.1.2, there is no certainty about why the asynchronous approach outperformed the synchronous approach on some dimensions of vocabulary knowledge. One plausible explanation is that the asynchronous approach reduces extraneous cognitive load, as hypothesised in section 2.5.3; equally, some other factor, such as increased time-on-task (Nation, 2007), may be at play. Follow-up studies could investigate the causal mechanisms involved by measuring whether cognitive load is actually lower during asynchronous reading-and-listening (Moussa-Inaty et al., 2012). Future research could also seek to control for time-on-task by comparing the synchronous and asynchronous approaches while keeping time-on-task constant. Such studies will face a dilemma: asynchronous reading-and-listening is inherently slower than synchronous reading-while-listening, because the written and aural input is consecutive rather than simultaneous; therefore, time-on-task can only be equalised by adding some other source of variation between the groups, such as giving the synchronous group extra material or asking them to read the story twice. Nevertheless, when considered alongside the present study, such research could help shed light on the mechanisms involved.

Finally, the impact of asynchronous reading-while-listening on other aspects of language learning should also be investigated. Future studies could explore learners' attitudes to the asynchronous approach, as well as its impact on comprehension and the development of reading and listening skills. Together, these avenues for future research would help build a picture of the potential of asynchronous reading-and-listening to enhance language learning.

Bibliography

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199-226. https://doi.org/10.1080/09588220802090246

Alotaibi, S., Pellicer-Sánchez, A., & Conklin, K. (2022). The effect of input modes and number of exposures on the learning of L2 binomials. *ITL-International Journal of Applied Linguistics*, *173*(1), 58-93. https://doi.org/10.1075/itl.21001.alo

Bisson, M. J., Van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language learning*, *64*(4), 855-877. https://doi.org/10.1111/lang.12085

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136-163.

Chang, A. C. S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, *37*(4), 652-663. https://doi.org/10.1016/j.system.2009.09.009

Chang, A. C. S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System*, *52*, 91-102. https://doi.org/10.1016/j.system.2015.05.003

Chang, A. C. S., & Read, J. (2013). Investigating the effects of multiple-choice listening test items in the oral versus written mode on L2 listeners' performance and perceptions. *System*, *41*(3), 575-586. https://doi.org/10.1016/j.system.2013.06.001

Chang, A. C., Wu, B. W. P., & Pang, J. C. (2013). Second language listening difficulties perceived by low-level learners. *Perceptual and motor skills*, *116*(2), 415-434. https://doi.org/10.2466/22.23.PMS.116.2.415-434

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied linguistics*, *31*(5), 693-713. https://doi.org/10.1093/applin/amq031

Chen, Y. (2021). Comparing incidental vocabulary learning from reading-only and Reading-while-Listening. *System*, *97*, 102442. https://doi.org/10.1016/j.system.2020.102442

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, *35*(1), 3-25. https://doi.org/10.1177/0265532216676851

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.), Lawrence Erlbaum Associates, Inc.

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257-276. https://doi.org/10.1177/0267658320921496

Darcy, I., & Fontaine, N. M. (2020). The Hoosier Vocal Emotions Corpus: A validated set of North American English pseudo-words for evaluating emotion processing. *Behavior Research Methods*, *52*, 901-917. https://doi.org/10.3758/s13428-019-01288-0

Davies, M. (2008). *The Corpus of Contemporary American English*. https://www.english-corpora.org/coca/

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and instruction*, *17*(1), 78-88. https://doi.org/10.1016/j.learninstruc.2006.11.007

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language learning*, *43*(4), 559-617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading research quarterly*, *24*(2), 174-187. https://doi.org/10.2307/747863

Gilman, R. A., & Moody, L. M. (1984). What practitioners say about listening: Research implications for the classroom. *Foreign language annals*, *17*(4), 331.

Haahr, M. (2023, August 7). *Random.org: True Random Number Service*. Random.org. https://www.random.org

Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The modern language journal*, *83*(3), 367-384. https://doi.org/10.1111/0026-7902.00028

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.

Kremmel, B. (2020) Measuring Vocabulary Learning Progress. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (Vol. 2). (pp.406-418). Routledge. https://doi.org/10.4324/9780429291586

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

Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. In B. Richards, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application*. (pp. 194-211). Springer. https://doi.org/10.1057/9780230242258

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners?. *Canadian Modern Language Review*, *63*(1), 127-147. https://doi.org/10.3138/cmlr.63.1.127

Moussa-Inaty, J., Ayres, P., & Sweller, J. (2012). Improving listening skills in English as a foreign language by reading rather than listening: A cognitive load perspective. *Applied Cognitive Psychology*, *26*(3), 391-402. https://doi.org/10.1002/acp.1840

Nation, I.S.P. (2005). Vocabulary Learning Though Extensive Reading. In G. Poedjosoedarmo (Ed.), *Innovative Approaches to Reading and Writing.* RELC Anthology Series 46. (pp.10-21). RELC, Singapore.

Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, *63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, *1*(1), 2-13. https://doi.org/10.2167/illt039.0

Nation, I. S. (2022). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/9781009093873

Nation, I. S., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart?. *Reading in a Foreign Language*, *22*(1), 31-55.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1-28.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484-503. https://doi.org/10.1017/S0261444812000018

Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading research quarterly*, *52*(1), 73-89. https://doi.org/10.1002/rrq.152

Teng, F. (2014). Incidental vocabulary learning by assessing frequency of word occurrence in a Graded Reader: Love or money. *LEARN Journal: Language Education and Acquisition Research Network*, *7*(2), 36-50.

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, *12*(3), 274-288. https://doi.org/10.1080/17501229.2016.1203328

Torgerson, D. & Torgerson, C. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. Springer. https://doi.org/10.1057/9780230583993

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language learning*, *69*(3), 559-599. https://doi.org/10.1111/lang.12343

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609-624. https://doi.org/10.1016/j.system.2013.07.012

Vu, D..V., & Peters, E. (2020). Learning Vocabulary from Reading-only, Reading-while-listening, and Reading with Textual Input Enhancement: Insights from Vietnamese EFL Learners. *RELC Journal*, 53(1), 85–100. https://doi.org/10.1177/0033688220911485

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

Waring, R., & Takaki, M. (2003). At What Rate Do Learners Learn and Retain New Vocabulary from Reading a Graded Reader?. *Reading in a Foreign Language*, *15*(2), 130-163.

Webb, S (2020). Incidental Vocabulary Learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (Vol. 2). (pp.225-239). Routledge. https://doi.org/10.4324/9780429291586

Webb, S., & Chang, A. C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68*(3), 267-290. https://doi.org/10.3138/cmlr.1204.1

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning?. *Language Teaching Research*, *19*(6), 667-686. https://doi.org/10.1177/1362168814559800

Webb, S., & Chang, A. C. S. (2022). How does mode of input affect the incidental learning of collocations?. *Studies in Second Language Acquisition*, *44*(1), 35-56. https://doi.org/10.1017/S0272263120000297

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

Wilkins, D. A. (1972). *Linguistics in language teaching*. The Chaucer Press.

Zhang, P., & Graham, S. (2020). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning. *Language Teaching Research*, *24*(6), 765-784. https://doi.org/10.1177/1362168819829022

Appendices

Appendix A: Texts

The three texts which the participants read and listened to are reproduced below. Note that these texts were not presented to participants in this form: the texts were used to make the videos which participants watched (Appendix B).

Text one: a small town and a storm

If you've ever lived in a small town, you know that it can be an interesting place sometimes. The town of Sleep City was not an interesting place. Around 400 people lived in or near the town. Most people from other places thought that the people in the town were yokels, but the people in the town didn't care if others thought they weren't smart. They mostly worked on farms – planting, taking care of their animals, and finding enough food for their families. It was a very simple life, and the people in the town liked it that way. They didn't talk very much, and they didn't like to listen to other people talk very much, either. They usually worked on their farms or stayed in their houses. When they walked into the town area, they usually went to the town store to talk with other people.

They built their houses away from others, on the tops of hills near big trees. That way, they could see if anyone was coming, and act like they were not at home. They loved their gentians and the other flowers that grew in the gardens. They didn't want anyone to walk on the grass around their homes. They were afraid of visitors to the town, and thought their stories of the city were prattle and just words. They didn't understand many of the things that the people from the city said. Nothing interesting ever happened in the town, until the storm came.

Since the people of the town didn't have TV or internet, they didn't know that the storm was coming. The people saw their bullocks and heifers running around in the field, but they didn't think anything of it. The strong wind blew the leaves in the sumacs and mastics, and it broke the flowers. That was only the beginning. The chickens and cats ran around on the fescue in the front yards because the walls of the animal buildings broke from the wind.

The storm got even worse. The wind blew harder, breaking many of the branches in the trees. When the walls of the animal buildings fell down, people saw their animals run away. The storm blew the gardens and flowers far away. The people didn't see any grass in their yards.

The oldest man in town came in to the store, where many of the people were waiting for the storm to pass. He said to the other yokels, "I've never seen a storm like this! I've lived in this town for 75 years, and there's never been a storm like it."

When the storm ended, the people came out of their houses to see what had happened. They found the heifers and the bullocks in the trees nearby. They took the branches from the mastics and sumacs out of their houses, and cut the wood to burn. They planted new fescue in their front yards, and it became green for the people to walk on again. The garden grew again with new gentians and other flowers.

Other people from the city came again to the town, and the people in the town still thought their words were prattle. It didn't seem like anything was different than before. The people in the town spent a lot of time talking about the storm after it was over, but otherwise it was the same. The very old man in the town wanted one change to his house, though. He bought a TV and got internet in his house. He wanted to know about the next storm *before* it happened, not *after*.

Text Two: the Good Man

There was once a city, ruled by a good king. His city was full of beautiful flowers, and there was peace in the land. The king loved animals, and had many animal friends. He loved to watch his dogs play, his animals work, and his birds sing. Some of his birds could fly to other lands to gather news, and he would send his kestrels and petrels to take news to other cities.

However, in the woods to the north there was a man. He was a bad man, and was leading a small cabal of people against the king. The bad man studied the dark spirits, and used flowers, morels and other plants to call the dark spirits.

The bad man used to be a good man, when he was the friend of the king. He would call on the good spirits using the beautiful freesia and flowers of the valley. He had a civet at that time and loved him very much. His animal friend would often go play with the dogs and sorels after their good work was done.

One day, his animal friend ate a bad morel from the forest and got very sick. The good man tried to call a spirit to help his animal friend live, but the king stopped the dabbler, and told him not to do it. The good man could not let his animal friend die, and went to the forest to call a spirit to heal the animal.

Two of the king's birds saw the good man in the forest and told the king. The King listened to the kestrel and petrel, and became angry. The King brought the Good Man back from the forest before he could call the spirit. Soon, the Good Man's animal friend died, and he was very angry. He left the dead animal near the river, and ran away from the city.

In the forest, he met other people who had also left the King, and he became the Bad Man. They decided to begin secret attacks against the King. The Bad Man and his friends took the plants from around the forest, and used dark spirits to kill the King's dogs. They also burned the beautiful freesia and other flowers in the fields.

The King was angry. He did not know where the dabbler had gone, and his birds could not find him. The Bad Man stayed in the forest, but was not happy with his new friends. He missed his animal friend. He decided he did not want to continue his attacks, and so he left the cabal.

As he walked in the woods, he found one of the King's birds with a broken wing. He helped the bird and asked him about the King. The Bad Man decided he would return to the King.

When The Bad Man returned, the King was not happy with him. However, the King told him he had something to give him if he helped to fix the problems he had caused. Then the King sent out his best sorels to find the Bad Man's friends.

After he helped the King find the group of people, the Bad Man became the Good Man again. To thank him, the King gave the Good Man a beautiful baby civet he had found. The animal had lost its family, so the Good Man took it into his house. Soon after, the King sent his birds to tell the other cities there were no more attacks. The land was at peace again, and the fields were again full of beautiful flowers.

Text Three: the Lawyer, the Old Man and the Ocean

One day, a ship left for a trip across the ocean with many people on it. People stood on the ship and watched spoonbills and other birds flying nearby. One man on the ship was a lawyer. Other people thought he was a shyster, who would try to get them to pay him a lot of money for him to be their lawyer. He was away from his job, and he enjoyed watching the sculpins and other fish.

Late one night, when the ship was very far out on the ocean, the ship hit a storm. It pushed the ship around on the ocean, right into a riptide. The ship turned over, and many of the people were thrown out of it. The storm continued all night, and the ship hit other riptides.

When morning came, the lawyer and an old man were the only ones alive. The big ship was in pieces on the top of the ocean. The two men lay on some of the pieces of the ship and tried to find food and water.

As they were lying on some of the jetsam from the ship, they saw birds flying near one area. They went to that area, and found some chervil and other plants and vegetables to eat.

The lawyer knew that the old man was the doyen on the ship, as he was many years older than the lawyer or any of the other people. The lawyer knew they would need to work together to stay alive. As they looked over the parts of the ship in the ocean, the men tried to find items to help them stay alive. The old man saw that the shyster found a long stick, so he used it to catch some fish.

The old man gave a happy burble as he caught the fish. They found some wood to burn to make a fire to cook the fish, and they cooked some of the chervil and other plants over the fire to eat.

The next day, the two men felt better, but they knew that they needed to find help. They saw some spoonbills in the sky, so they knew that there must be islands nearby. Then, the lawyer saw the islands far away. The old man wanted to help the lawyer look through the jetsam of the ship to try to find something to move their small boat in the direction of the islands. They began looking through the broken ship. As they looked for something to help them, the lawyer heard the doyen shout that he had found a radio to call for help. His shout frightened the birds in the sky and the sculpins under the water, and the old man saw the lawyer almost fall off his boat in surprise. He said "We might be saved after all!"

The two men used the radio to call for help. They talked and talked into the radio for hours. When they finally heard a happy burble on the other side of the radio, they knew they would be safe. The birds looking down saw the lawyer and the old man dancing and singing. Some people from the islands came to help them, and they were able to return home.

When he got back home, the lawyer changed his ways and became good. He worked hard as a lawyer, and people liked him. The old man spent more time with his family, as he knew that he wouldn't live for very long. Their time on the ship and in the ocean made them very good friends.

Appendix B: Videos

Test videos

Synchronous: https://youtu.be/e_16CigHVqo

Asynchronous: https://youtu.be/5bUJFwvT7Xo

Video one: A small town and a storm

Synchronous: https://youtu.be/6MjG7aFVaEU

Asynchronous: https://youtu.be/OG2jBTj10HY

Video two: The Good Man

Synchronous: https://youtu.be/jydJlGF-ndM

Asynchronous: https://youtu.be/U0KtCSpo9k8

Video three: The Lawyer, the Old Man and the Ocean

Synchronous: https://youtu.be/wJzE-cwPKnk

Asynchronous: https://youtu.be/chm3phcqc9A

Appendix C: Comprehension Questions

Video one: A small Town and a Storm

1. What was life like for people in the town?

A. Simple B. complicated

2. What did people from the town think of visitors?

A. They liked them B. They were afraid of them

3. During the storm, what happened to the walls of the animal buildings?

A. They fell over B. They stayed standing

4. What did the very old man do after the storm?

A. He built a new house B. He got internet and TV in his house

Video two: The Good Man

5. Where were the dark woods?

A. To the South B. To the North

6. Did the good man try to heal his animal friend?

A. Yes B. No

7. What was wrong with the King's bird in the forest?

A. It had a broken wing B. It was very tired

8. At the end of the story, what did the land have?

A. Peace B. War between different countries.

Video three: The Lawyer, the Old Man, and the Ocean

9. How many people from the ship were alive after the storm?

A. Two B. Five

10. How did the two men cook the food?

A. In the kitchen B. Over a fire

11. What did the men find from the trip?

A. A TV B. A radio

12. Did they men change the way they lived their lives?

A. Yes B. No

Appendix D: Immediate form-meaning test

*Please select the correct meaning for each word from the stories. Do **not** look the words up - just answer from your memory.*

1. A shyster is

A. a person B. a trip C. a door

2. A heifer is

A. a university B. a sign C. an animal

3. A doyen is

A. a fish B. a bag C. a person

4. A spoonbill is

A. a kind of hair B. a bird C. a mirror

5. Fescue is

A. grass B. gum C. paint

6. Chervil is

A. something you watch B. something you eat C. something you wear

7. a riptide is

A. on an island B. in an ocean C. in a forest

8. A mastic is

A. a tree B. a chair C. a bus

9. a petrel is

A. a hat B. a store C. a bird

10. Freesia are

A. clothes B. flowers C. signs

11. a kestrel is

A. a bird B. a wall C. a restaurant

12. A yokel is

A. a bottle B. a kind of food C. a person

13. a sorel is

A. a kind of table B. an animal C. a shape

14. A gentian is

A. a computer B. a flower C. a book

15. a bullock is

A. an animal B. a kind of house C. a kind of airplane

16. A sculpin is

A. a cord B. a fish C. a letter

17. Jetsam is

A. on the land B. on the ocean C. in the air

18. A morel is

A. a place B. a plant C. a person

19. a dabbler is

A. a type of animal B. a type of plant C. a type of person

20. Burble is

A. a person talking B. a light shining C. the wind blowing

21. Prattle is

A. lots of food B. lots of words C. lots of books

22. a sumac is

A. a tree B. a town C. a trip

23. A cabal is

A. a group of people B. a type of bird C. a type of food

24. a civet is

A. a bank B. an animal C. a camera

Appendix E: Immediate Form-Recognition Test

Please look at the list of words below and select every word that you remember seeing (or hearing) in the videos.

It doesn't matter what the words mean. Just select every word that was in the videos.

- nervack
- sorel
- shyster
- heifer
- voker
- ziddy
- civet
- domner
- spoonbill
- bullock
- dabbler
- lairet
- tannock
- jetsom
- fescue
- volers
- vigging
- cabal
- armidge
- burish
- sculpin
- nipher
- gentian
- infess
- kervo
- petrel
- larpy
- burble
- yokel
- modner
- kestrel
- merrus
- nidder
- chervil
- affning
- prattle
- morel
- shifin
- doyen
- sumac

- slover
- terril
- freesia
- ramidge
- riptide
- thager
- mastic
- nillen

Appendix F: Delayed form-meaning test

[Note: the delayed form-meaning test was identical to the immediate form-meaning test]

*Please select the correct meaning for each word from the stories. Do **not** look the words up - just answer from your memory.*

1. A shyster is

A. a person B. a trip C. a door

2. A heifer is

A. a university B. a sign C. an animal

3. A doyen is

A. a fish B. a bag C. a person

4. A spoonbill is

A. a kind of hair B. a bird C. a mirror

5. Fescue is

A. grass B. gum C. paint

6. Chervil is

A. something you watch B. something you eat C. something you wear

7. a riptide is

A. on an island B. in an ocean C. in a forest

8. A mastic is

A. a tree B. a chair C. a bus

9. a petrel is

A. a hat B. a store C. a bird

10. Freesia are

A. clothes B. flowers C. signs

11. a kestrel is

A. a bird B. a wall C. a restaurant

12. A yokel is

A. a bottle B. a kind of food C. a person

13. a sorel is

A. a kind of table B. an animal C. a shape

14. A gentian is

A. a computer B. a flower C. a book

15. a bullock is

A. an animal B. a kind of house C. a kind of airplane

16. A sculpin is

A. a cord B. a fish C. a letter

17. Jetsam is

A. on the land B. on the ocean C. in the air

18. A morel is

A. a place B. a plant C. a person

19. a dabbler is

A. a type of animal B. a type of plant C. a type of person

20. Burble is

A. a person talking B. a light shining C. the wind blowing

21. Prattle is

A. lots of food B. lots of words C. lots of books

22. a sumac is

A. a tree B. a town C. a trip

23. A cabal is

A. a group of people B. a type of bird C. a type of food

24. a civet is

A. a bank B. an animal C. a camera

Appendix G: Delayed form-recognition test

Please look at the list of words below and select every word that you remember seeing (or hearing) in the videos last week.

It doesn't matter what the words mean. Just try to select every word that was in the videos.

- menno
- sorel
- shyster
- heifer
- joules
- lowmen
- civet
- lebby
- spoonbill
- bullock
- dabbler
- youssle
- zilard
- jetsom
- fescue
- lorack
- nammy
- cabal
- valish
- tairack
- sculpin
- agerth
- gentian
- revo
- leknodge
- petrel
- vashil
- burble
- yokel
- scopies
- kestrel
- mowan
- fingill
- chervil
- winnith
- prattle
- morel

- elby
- doyen
- sumac
- dernom
- madage
- freesia
- nomel
- riptide
- vishal
- mastic
- zuber

Appendix H: Webpages

This appendix contains links to the webpages which were sent to participants. These webpages bring together the tests, videos and instructions into one place.

Synchronous: https://forms.gle/doWehjNFxqZJuacc7

Asynchronous: https://forms.gle/qCRT8K4EAyi6szkY8

Delayed test: https://forms.gle/SjaXPeTcnMXn74yG7

Appendix I: Participant Information Sheet

PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: [redacted]

You are being invited to take part in a research project. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

1. **Why is this research being conducted?**

This research is for a master's dissertation. The masters is in applied linguistics, which is the study of how people learn languages. The research looks at the effects of reading to a text while listening to it at the same time. I want to understand how reading while listening can help learners like you improve their English.

2. **Why have I been invited to take part?**

Your English teacher told you about my research, and you said you would like some more information to decide if you want to take part. You have been approached because you are an adult English student and your level is approximately intermediate or higher. In total, I plan to recruit about 40 participants.

3. **Do I have to take part?**

No. It is up to you to decide whether to take part. If you want to take part, please fill in the consent form which I will send you.

You are free to change your mind and withdraw yourself from the research, without giving a reason, and without negative consequences, by telling me that you want to withdraw.

If you want to withdraw your information after taking part in the research, you can do so before July 1, 2023, by contacting me and telling me that you want to withdraw. I will then delete all the information you have given me. After July 1, it will not be possible to withdraw your information.

4. **What will happen to me if I take part in the research?**

All the research will be done online.

First, you will fill in an online form to give your consent to join the research and provide some information about yourself (for example: your age).

Then, you will join a research session via an online video call. We will schedule the call at a time that is convenient for you. I will join the video call to give you the information you need and help you with any problems. There may be a small group of other participants joining you on the call. If you don't want to turn on your video, you don't have to – audio (voice) calls only are fine.

The call will last about one hour. You will read and listen to three short stories. Then, you will take some short tests.

After this, you will have the option to join a second call one week later, for about 15 minutes. In this call, you will take some short follow-up tests. This second call is optional – you don't have to join it.

You do not have to worry about the tests: they are not designed to evaluate you personally; they are designed to evaluate the teaching methods used in the research. You will be free to ask to pause or stop the research activities at any time.

5. **What are the possible disadvantages and risks in taking part?**

There are not expected to be any disadvantages or risks in taking part. Your data will be anonymised (your name will be removed) before publishing the results.

6. **Are there any benefits in taking part?**

Taking part in the research will give you some practice in reading and listening in English.

You will also have the opportunity to join a free online lesson delivered by the researcher on a topic chosen by participants, if you wish.

7. **Payment**

All participants successfully complete the first video call will have the opportunity to enter a prize draw. In the prize draw, four participants will be chosen at random and will win $50 each in vouchers of your choice (there will be a total of about 40 participants). If you do not want to enter the prize draw, please tell me and I will remove your name.

8. **What information will be collected and why is the collection of this information relevant for achieving the research objectives?**

The following information will be collected about you: your name and contact details; your consent forms; a questionnaire about some personal details such as your age; the time and day that you took part in the research; the results of the tests you take during the research.

Identifiable data (including consent forms) will be on stored on my password-protected Oxford University OneDrive account. It will be deleted when the research is completed and my dissertation is submitted, in August, 2023. While I am doing the research, I may also store your contact details on my password-protected phone or in my password-protected email account so that I can contact you to administer the research; this data will be deleted once you have finished the second session of the research.

Other research data will be stored for at least 3 years after publication or public release of the work of the research. It will also be stored on my password-protected Oxford University OneDrive account. The researcher, Chris Lack, and my supervisor, Hamish Chalmers, will have access to the research data.

Research data may be transferred to, and stored at, a destination outside the UK and the European Economic Area. Any data transfer will be done securely and with a similar level of data protection as required under UK law.

9. **Will the research be published? Could I be identified from any publications or other research outputs?**

The findings from the research will be written up in a dissertation. A copy of my dissertation will be deposited both in print and online in the Oxford University Research Archive where it will be publicly available to facilitate its use in future research. It may also be written up in an academic publication or conference presentation, or on a website. It will be not possible for participants to be identified from the outputs.

10. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance web site at https://compliance.admin.ox.ac.uk/individual-rights.

11. Who has reviewed this research?

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: [redacted]

12. Who do I contact if I have a concern about the research or I wish to complain?

If you have a concern about any aspect of this research, please contact [redacted] or [redacted] and we will do our best to answer your qu [redacted] ys and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

The Chair, Medical Sciences Interdivisional Research Ethics Committee;
Email: ethics@medsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

The Chair, [Education Research Ethics Committee

Email: student.curec@education.ox.ac.uk Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

13. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

[redacted]

Appendix J: Participant Consent form

Consent to take part in a study on the benefits of Reading while Listening to English texts

Central
University Research Ethics Committee (CUREC) approval reference: [redacted]
Purpose of study: to understand how reading while listening can help learners improve their English

* Indicates required question

1. I confirm that I have read and understand the information sheet for this research. I have had the opportunity to think about the information, ask questions and have had these answered satisfactorily. *

   Mark only one oval.

   ◯ I agree

2. *

   I understand that my participation is voluntary and that I am free to withdraw at any point until July 1, 2023, without giving any reason.

   Mark only one oval.

   ◯ I agree

3. *

   I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.

   Mark only one oval.

   ◯ I agree

4. I understand that I will not be identifiable from any publications or other research outputs. *

Mark only one oval.

◯ I agree

5. I understand how to raise a concern or make a complaint. *

Mark only one oval.

◯ I agree

6. I agree to take part. I understand that by completing this page and clicking "next", I am consenting to join the study. *

Check all that apply.

☐ I agree

7. What is your full name? *

Appendix K: Background Information Survey

Information about you

Please give some information about yourself by answering the questions below

8. What is your date of birth? *

Example: January 7, 2019

9. What is your first (native) language?

10. What is the highest level of degree or school you have completed?

Mark only one oval.

- ◯ I have never attended school
- ◯ primary school (up to age 11)
- ◯ High school
- ◯ Undergraduate degree
- ◯ postgraduate degree (eg: masters', doctorate)

11. For how many years have you been studying English?

12. What level is your English?

Mark only one oval.

- ◯ Intermediate
- ◯ Upper Intermediate
- ◯ Advanced
- ◯ Other:

13. What is your gender?

Mark only one oval.

- ◯ Male
- ◯ Female
- ◯ Other
- ◯ Prefer not to say

Appendix L: Interview questions for pilot participants

Technical issues:

Did you have any technical problems?

Were the instructions clear while you were doing the tasks?

Was it easy to navigate between the form and the videos?

Understanding the videos

Did you understand the story in the videos? How easy/difficult was it to understand?

How was the speed of the audio? [prompts: too fast, too slow, about right?]

Did you have the right amount of time to read the text properly? [prompts: too much time? Not enough time?]

[for the synchronous group:] How did you find the format of reading and listening at the same time? [prompts: was it helpful/clear/confusing/distracting?]

[for the asynchronous group]: How did you find the format of reading a small part of the text then listening to it? [prompts: was it helpful/clear/confusing/distracting?]

Tests:

Did you have any problems while taking the vocabulary tests?

Did you have any problems answering the comprehension questions?

Think about the vocabulary tests. Did you know (or had you ever encountered) any of these words before today? Which ones?

General:

Did you have any other problems or suggestions? Is there anything I should know about to make these materials better?

Appendix M: Test results

Partc-ipant num-ber	Experimental Group	Comprehension Test Results	Immediate Form Recognition Test Results	Immediate Form-Meaning Test Results	Delayed Form Recognition Test	Delayed Form-Meaning Test Results
1	Asynchronous	12	20	21	20	22
2	Asynchronous	11	18	21	19	21
3	Asynchronous	12	19	22		
4	Asynchronous	12	5	18	10	17
5	Asynchronous	11	17	19	10	15
6	Asynchronous	12	8	22	-3	17
7	Asynchronous	11	15	18	10	15
8	Asynchronous	12	3	14	2	13
9	Asynchronous	11	10	15	7	7
10	Asynchronous	12	5	18	3	14
11	Asynchronous	10	9	12	4	9
12	Asynchronous	12	9	20	7	22
13	Asynchronous	11	7	16	8	14
14	Asynchronous	11	21	20	19	22
15	Synchronous	12	16	19	17	19
16	Synchronous	12	14	16		
17	Synchronous	11	14	19	10	19
18	Synchronous	12	20	20	16	21
19	Synchronous	12	13	19	19	19
20	Synchronous	11	14			
21	Synchronous	11	5	16	5	13
22	Synchronous	12	1	21		
23	Synchronous	11	4	15		
24	Synchronous	12	4	13	1	17
25	Synchronous	10	-2	11		
26	Synchronous	8	6	11	7	10
27	Synchronous	11	7	9		

Appendix N: CUREC approval email

From: [redacted]
Date: Tuesday, 14 March 2023 at 20:11
To: [redacted]
Cc: [redacted]
Subject: Re: New CUREC application [redacted]
Dear [redacted]

Research ethics approval

Research title: Incidental Vocabulary Acquisition from synchronous vs asynchronous reading-while-listening

Research ethics reference: [redacted]

The above application has been on behalf of the Education Departmental Research Ethics Committee (DREC) in accordance with the University's procedures for ethical approval of all research involving human participants.

I am pleased to confirm that, on the basis of the information provided to the DREC, ethics approval has now been granted for this study. Please find approval letter attached.

Please note the following:

Personal data: It is the responsibility of the PI to ensure that all personal data collected during the project is managed in accordance with the University's guidance and legal requirements.

In-person activities: Any data collection involving in-person interactions with participants must have an up-to-date fieldwork risk assessment in place; further guidance is available from the Safety Office's website.

Amendments: Please notify the committee if you intend to make any amendments to the information in your ethics application as submitted at date of this approval, as all changes must receive ethical approval prior to implementation. The amendment form is available on the SSH IDREC webpage.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to staff.curec@education.ox.ac.uk / student.curec@education.ox.ac.uk or ethics@socsci.ox.ac.uk .

Yours sincerely
[redacted]