An investigation to see whether listening whilst reading helps immediate comprehension compared with just reading in L2 English learners.

Jade Greenway

MSc in Applied Linguistics and Second Language Acquisition, 2024

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]:	
Date:	09/08/2024

Abstract

Text-to-speech software has consistently been shown to help students with learning disabilities, yet the benefit of reading-whilst-listening for wider populations has been under-researched. One specific population which may benefit from this audio addition is second language learners. Therefore, this study investigated whether reading-whilst-listening can help a wide demographic of English language learners. A counterbalanced reading comprehension study was completed with participants having to read two texts, one with audio and one without. Their scores on comprehension tests for these two texts were then statistically analysed to see if audio made a difference. The results showed for one of the texts, the addition of audio when reading led to higher comprehension scores whilst in the other text the addition of audio hindered comprehension. Participants chose their own audio speed from three options based upon which they preferred, and exploratory analyses showed that there was no evidence that audio speed affected later comprehension performance. These findings are both in contrast to results from previous similar studies with additional English language learners which found null results and the results are also in alignment to previous results due to the second text. These results are discussed in terms of the larger sample size and wider additional language learners demographic in the current study. Thus, further research is needed.

Key words: text-to-speech, reading-whilst-listening, additional language learning, English learners, reading comprehension

Table of Contents

Acknowledgements

I would firstly like to thank my supervisor, Elizabeth Wonnacott. Liz, thank you for all your help and support. Thank you for keeping me calm and answering all my questions.

I would also like to thank Holly Jenkins for helping me so much with my experiment and gathering data.

I would like to thank all my participants; I would not have been able to complete this without them.

Thank you to my friends, the ones from home and the ones I have gained this year. You have all kept me going, kept me laughing, and kept me working.

A special thanks goes to Linda Somers, my dyslexia tutor. Without your support and confidence, I do not know if I would have finished this. It was a real struggle this time. You have been there for me over the last four years, and I have gained a friend through your support. Without you helping me read articles I would still be staring at a screen!

Finally, a big thank you to my number one supporter, my dad. Without you I never would have been able to come here and realise my potential. You have dealt with so much from me, but you have been there throughout, never wavering in your love and support. So, thank you dad, for everything you do for me.

List of tables and figures

List of abbreviations

Mdn = median

RQ = research question

RQ1 = research question 1

RQ2 = research question 2

RQ3 = research question 3

Gorilla Experiment Builder = Gorilla.sc

Unique identification = unique id

1. Introduction

One of the main skills needed when acquiring knowledge of a language is the ability to read. This is a challenge learners tend to face when dealing with an additional language due to the complex nature of reading and has therefore led to an area of research that investigates techniques that can be used to help learners. This ability makes it possible to learn vocabulary and grammar from exposure to written texts, and therefore any supportive techniques are beneficial. However, difficulties in reading are not solely found in second language learners; students who have a learning disability also often have a difficulty with reading large bodies of text. Therefore, some scholars (Zhang & Zou, 2022) have tried utilising techniques that help students with dyslexia, to see if those same techniques can help students with learning an additional language. One such technique that has been researched is reading-whilst-listening, which is the technique chosen for this study due to the positive impact it has on students with learning difficulties. Reading-whilst-listening has been used by researchers to investigate various areas of literature, including phonological awareness, vocabulary acquisition and language awareness (Borrás & Lafayette, 2011; Hempenstall, 1997; Woodall, 2010). Asrimawati and Margana (2020), for example, investigated this same area; their analysis study of 10 previous papers, all of which had a small sample size, considered whether there was a specific benefit to reading-whilst-listening in additional language classrooms. The findings of these papers were mixed; however, there were marginally more findings to suggest that, over time, reading-whilst-listening had a positive benefit for learners.

The reason for this current investigation was curiosity as to whether the tools I use to help my own dyslexia could also be helpful for second language learners. Other research has found that interventions originally intended for individuals with disabilities are also useful for L2 learners. For example, it was found that subtitles can help additional language learners (Birulés-Muntané, & Soto-Faraco, 2016; Vanderplank, 2016) and therefore, I was interested to see if the same can be found for other aspects of technology.

Previous research (Draffan et al., 2007; Gotesman & Goldfus, 2010) has investigated whether additional language learners can benefit from additional audio when reading. Former investigations have led to mixed results in this area, with some having positive results and some null. However, the studies either had a small sample size or the participants were gathered from one place, and consequently, generalisations could not be made. Additionally, a main concern is the contradictions in findings as there are mixed results. One of the aims for this sample was that there would be more

participants from a wider demographic background to allow for more diverse representative results. The aim of this dissertation was to investigate whether text-to-speech software can be utilised when reading-whilst-listening to help second language learners' reading comprehension when comparing comprehension scores to just reading. The purpose is to see if text-to-speech can be a viable technological help for additional language learners to comprehend what they are reading at a higher level. 45 English language learners participated in this online counterbalanced study. All participants were of intermediate language level and gained from multiple streams of promotion.

Some scholars, like Krashen (2013) and Bamford and Day (1998), believe that usage of this skill can help create proficient second language users due to the amount of vocabulary and grammatical knowledge acquired. One limitation second language learners have though, in comparison to first language learners, is that they are utilising the skill of reading at the same time as they are learning the oral language. Therefore, adult learners may benefit more from aural languages, and the addition of audio of texts more when reading compared with children due to their learning order. Verhoeven (1990) emphasised the difficulty children face and explained how "reading instruction should be matched to those [oral] skills" (p.90) because of the benefits reading can have on language acquisition rather than students trying to read ahead of their level. This is also an important point because it demonstrates why the addition of audio could be beneficial when reading as it allows for the texts to be pitched to the correct level and allows for advancement in both skills.

In this dissertation, this first section is the introduction in which I have outlined the key aims and objectives. In the next section I will firstly look at previous literature, to find and explain the research gap, before designing the research questions for this study. The third section will explain the methodology employed and explain the demographic of the participants. The results will follow in the next sections, showing what the statistical tests found before the discussion of what was found is presented in section five. Finally, a conclusion will explain all the main points and reiterate the findings whilst providing ideas for future research and explaining any possible limitations of this experiment.

2. Literature Review

2.1 Introduction

Stein et al. (2011) explain that spelling and reading are lifelong abilities that improve quality of life. Therefore, when learning languages, reading and understanding correctly are key parts of language development. One way this is taught in classrooms is through reading comprehensions. In this thesis, the definition of reading comprehension used is taken from Koda (2005) who describes reading comprehension as a skill where "the reader extracts and integrates various information from the text and combines it with what is already known" (Koda, 2005, p.4).

In this section, I will explain and evaluate previous literature with the aim of contextualising this experiment and demonstrating how it fits into a research gap. This will be done using thematic analysis. Four main theme groups were identified. Two groups highlight typically developing students in their first and then in their second languages, whilst the other two focus on dyslexic participants in their first and then in their second languages. Finally, I will summarise where this experiment fits within the research gap identified by this literature review before identifying the research questions and hypotheses.

2.2 Definitions

This experiment is focused on two modalities, reading-whilst-listening and reading only, and whether the addition of audio when reading can have a benefit on reading comprehension. The research deliberated over in this section will be focussed on these two modalities in participants' first and second languages. The research behind this investigation was created due to the benefit reading-whilst-listening has for dyslexics. Dyslexia is a learning disability which impacts students in multiple different ways, but one key component is difficulties in word recognition. One technique that is used by students with dyslexia to help lessen their load is to use text-to-speech software. This technique has been researched over the years, proving time and time again to be effective (Bhola, 2022; Dawson et al., 2019; 2024 ,الفل & ابراهيم).

One interesting finding by Liao et al. (2020) was that subtitles, originally created to help deaf people watch television, also helped students to learn additional languages. Therefore, I wanted to investigate whether text-to-speech software, used by dyslexics for reading-whilst-listening, could benefit additional language learners too. Previous literature, as will be discussed in this section, led to mixed results and limitations, therefore the aim of this research is to fill the gap and provide more

information on this software being used when reading-whilst-listening. Therefore, some theoretical ideas as to why text-to-speech, and reading-whilst-listening in general, might be important for language learners will be addressed. One idea is that as text-to-speech software can help dyslexics, it may be useful for language learners too.

In this experiment, reading-whilst-listening is identified using Koh's (2024) definition, where someone is “simultaneously reading and listening to the oral renditions of a written text” (p.163). Reading-while-listening can be performed in different ways: with teachers reading to the student (Albright & Ariail, 2005; Beck & McKeown, 1991; Hoffman et al., 1993); audio tapes being used, including CDs in classrooms (Verlaan & Ortlieb, 2012); audio books (Chang, 2011) or online extensions like text-to-speech applications (Roberts et al., 2013).

This style of reading has been used by students for many years, with researchers discussing the benefit in first languages (Aoki et al., 2022; Handley, 2009) and in additional languages too (Drezek, 2001; Oktalia & Drajati, 2018). The most commonly mentioned benefit is in reading comprehension due to the cognitive load burden being lessened (Govender & King, 2018). Whilst researchers have also found other positive effects of reading-whilst-listening, including reading rate and vocabulary knowledge (Grunér et al., 2018; Harvey et al., 2013; Syrdal et al., 2012), reading comprehension will be the focus of this investigation. This focus is particularly important for research because results are not conclusive on the benefit reading-whilst-listening has on comprehension.

In this experiment, the assumption was made that the reading only condition is reading silently in one’s mind. This phenomenon is a natural skill we partake in daily (Turker et al., 2024). Therefore, it is a common occurrence in both first and second language classrooms which teachers ask students to partake in (Rupley et al., 2020). Reading only will be the other modality in this experiment due it being commonly used in education and will therefore be used as a control. One thing to note about reading silently is that researchers do not agree on how this phenomenon takes place. Some researchers, as shown by Langland-Hassan (2021) believe a voice is ‘heard’ whilst reading occurs. This voice is often referred to as inner speech and researchers have shown there are similarities to overt speech. However, there are some researchers who argue that the differences diminish the similarities (MacKay, 1992; Nooteboom, 2010). Due to the differing results in literature, it is less clear how second language learners read silently. Bassetti (2008) emphasised how additional languages typically have stronger orthographical than phonological representations. Therefore, it will also be interesting to see if giving participants the chance to not rely solely on silent reading helps different populations.

Rubin (1994) performed a systematic review of "second language listening comprehension research" (p.199) where one of his research questions looked into when and how listening is performed to gain comprehension. Their paper reviewed 130 papers, with 115 being deemed adequate for inclusion. These papers resulted in the conclusion that the focus should be on the differences that are present in listening comprehension and not the similarities. For example, each proficiency level has different needs and skills that need addressing. In particular, making the audios natural and engaging for listeners, so they can negotiate meaningful information, is essential. One way that this can be accomplished is by finding the most beneficial aspects of audio for listeners, for example different speeds, accents or colloquial phrases.

2.3 Studies with typically developing participants in their first language

Amer (1997) tested whether having the teacher read the text aloud to the students benefited the student comprehension. There were 75 students taking part in the experiment, who were split into two groups with the control group reading silently and the experimental group reading-whilst-listening with a teacher reading out to them. They then had to answer 15 multiple choice questions and fill in the blanks in a story frame. Their results showed that there was a positive impact to having the teacher read to the students as they read along meaning that the human voice alongside helped.

Dai et al. (2022) investigated how text-to-speech is used in education for 17-18-year-olds, and similarly had a human voice read. 44 Dutch students were asked to listen to five Dutch audio fragments; therefore, all results were based on reading-whilst-listening. In their study, they created four different audio texts using different amounts of training data to see if different voices have an impact on the comprehensibility of audio. The training data was based on the amount of time the software had been fed desired speaker characteristics. This within-subject experiment had five different audios (one human and four created) read five different texts. A positive of this experiment is that they changed the audio speed to 0.7 times slower than normal speed, so children had an easier time of reading and listening at the same time, however, there was no indication of why this specific speed was chosen after the pilot experiments. The aim of this experiment was to see how much training synthetic voices needed for students to comprehend and recall what they read. Another limitation to their study was that they had no group that did not listen to the audio, and therefore the benefits of the audio were only related to other audios. The results of this experiment were interesting because the researchers claimed that text-to-speech was a useful feature for

students to help with comprehension, however, their findings only focused on audio and showed that the human voice was more beneficial than the text-to-speech software. Additionally, the results demonstrated a plateau effect with the amount of training, meaning that more input did not necessarily improve the audio output for the students. Therefore, this study demonstrated that the more natural voice was the most well received audio, but none of the text-to-speech audios led to a specifically negative comprehension score.

Lionetti and Cole (2004), on the other hand, investigated the speed rate of the audio to see if the speed matters in providing students with a benefit when reading-whilst-listening. Four children from a private school in Pennsylvania took part where they listened to an audio recording of the investigator reading texts at two different speeds. They predicted that there would be a significant difference, in the three test scores. The first and second tests were performed when the participants were reading the passage again after hearing it once. They were tested on the words spoken and the accuracy. The final test was a comprehension test out of 100. Their results showed that neither speed was particularly favourable for reading comprehension.

Mayer (2002) explains why multimedia is vital in education. This theory of education proposes that meaningful learning is completed by the learner, and that they need to create the connections in their brain between words and images/sounds for learning to occur. In his article, Mayer explained multiple principals that he has developed from his research. Dousay (2016) expanded on this idea and found that the impact of multimedia learning can be nuanced and is more context-dependent due to the differing effects. In their experiment, 102 participants were split into three groups: animation-text, animation-narration, and animation-narration-text, to see which group was the most interested in the texts. A “situational interest survey” (Dousay, 2016. p.1263) was administrated at the end of the experiment to gain the results. The findings showed that there can be impact on motivation if provided different tools for learning, but it depends on the learner. This finding shows that any additions in learning can be helpful.

Parr (2012) investigates how students felt about text-to-speech software and how it is all person-dependent. In his investigation, 28 typically developing students were questioned on their views of text-to-speech before, during and after adding the software into their education over the course of eight months. It was found that students need the time to understand how the software works and that it should be viewed as a helper, not the final product.

One supposed and researched benefit of reading as a skill is incidental learning. There is evidence that reading, alone or whilst listening to audio, can help both first and second language learners

improve their knowledge of vocabulary by them comprehending the meaning behind what they are reading (Chen, 2021; Daskalovska, 2014; Mohamed, 2018; Paribakht & Wesche, 1999; Waring & Nation, 2004; Yali, 2010; Zhao et al., 2016). For example, Zhao et al. (2016) found that students with a higher proficiency in the target language, picked up more words without conscious effort when they were reading compared to participants with a low proficiency. In their experiment, 129 English learners from a university in China participated in two surprise comprehension and vocabulary tests after reading to see how much they learnt incidentally. A total of 20 target words were chosen for the surprise test from the two texts the participants had read, to see if there are correlations between three different factors (motivation, anxiety and strategy knowledge) and the amount of incidental learning. All three aspects were found to correlate, demonstrating that reading is an effective learning method.

2.4 Studies with typically developing participants in their second language

Reading comprehension in additional languages is a heavily researched area of study with there being multiple factors at play influencing the acquisition. For example, reading comprehension can be influenced by vocabulary knowledge (Zhang & Zhang, 2022), pedagogical practices (Day et al., 1998), sociocultural contexts including assistive technology, and student abilities.

One of the most heavily researched areas of reading comprehension is vocabulary knowledge because knowing vocabulary is fundamental to understanding a language. Nation (2005) makes this importance clear when he explains the different elements of learning a language's vocabulary. "The number of words in the language, the number of words known by native speakers and the number of words needed to use the language" (p.6) are all aspects that are necessary for researchers and teachers to know for correct understanding. Nation's assessments are widely accepted and have led to more research into each of these areas.

Cho and Hey-Jung (2004) investigated the benefit of recreational reading in the target language; in this research, 140 English language learners in Korea took part in an experiment to see if there was a benefit to reading in language learning alongside traditional education. Their results showed that the 70 students in the experimental group who partook in additional reading on top of their studies had superior gains across all areas of language studied compared to the 70 participants who only had traditional classes.

Furthermore, it has also been found that when reading, having a longer text to read rather than just short sentences is more beneficial for learners (Ktistakis et al., 2024). This is an interesting concept to remember due to the English reading tests learners perform. Typically, as shown by Wissinger et al. (2024), reading tests often are brief and focus more on testing short sentences rather than more detailed texts. Therefore, if there is even more of a benefit for learners when reading longer passages, reading comprehension tests should be focussed on general meaning and understanding as well as short sentences.

Another area of research in second language reading comprehension is the pedagogical affects. The way we teach and educate learners is important and new strategies are constantly being created (Carrell et al., 1988). In this section, a few articles that look at some of these strategies will be looked at, however, one thing to remember is that using text-to-speech in the classroom is a new pedagogical practice in itself.

Understanding why reading-whilst-listening might be useful is also an important area of study. One of the main theoretical frameworks when discussing comprehension is the cognitive load theory, first proposed by Sweller (1988) and further discussed by Plass et al. (2010). The idea behind this theory is that the human brain can only work and remember so much, before the cognitive load is reached. In education, effective materials and additional support are believed to help expand and lessen the strain of the cognitive load. The idea is that for new skills and information to be learnt to a level students can use, they must be added to a student's long-term memory and therefore we have a limit on what we can remember.

Reading-whilst-listening may reduce cognitive load and is important for multimodal learning, however there are mixed views in literature. Kim (2020), for example, investigated whether reading-whilst-listening helps more difficult passages seem easier. Their findings showed that, of the five participants, none of them believed the text-to-speech to be helpful despite the results showing three having slightly better scores.

Mayer and Moreno (2003), contrastingly, support the idea that multimedia learning helps ease cognitive load by providing multiple examples and responses to different situations. They suggest, and explain, how presenting information through multiple sensory channels can help distribute the extent of cognitive load and allow for more language acquisition. In this article, they explain different scenarios of cognitive overload, and how multimedia learning could help to lower this burden. For example, one type of overload they explain is when our mind becomes overloaded with "essential processing demands" (p.45), in this example, they describe how providing a different input

for the information allows the brain to comprehend two different strands of the same information, reducing stress.

Kiamanesh et al. (2023) examined two different areas of text-to-speech technology: the modality as in the experiment reported in this thesis, and the accent. Their first experiment is the more relevant for this paper; however, their second experiment led to some interesting findings.

In their first experiment, and the experiment that is more relevant for this thesis, Kiamanesh et al. investigated the ways in which reading text (whether reading alone or reading-whilst-listening, or just listening to a text), can impact comprehension and working memory. 157 participants participated in experiment 1, with participants randomly allocated to one of the three groups. For this exercise, one text about the Yellowstone national park was chosen with 20 comprehension questions following the text. Kiamanesh et al. showed the text to the students one paragraph at a time in the reading section with the length of time matching the length of the audio. Once the participants had seen the material, they predicted how much they understood before answering the question. Based on their aim "to collect around 50 participants per condition" (p.9) we can assume the participants were roughly split around this number in the three groups. The results showed that there was no significant difference between the text-only and dual modality groups in their comprehension scores, although participants in the text-only group did predict higher scores when asked how much they believe they remember. When looking at the scores themselves, participants in the text-only group had numerically higher scores than the participants in the other two groups, however, this did not lead to a significant difference, and may be due to the participants themselves as no pre-tests were performed to gain initial knowledge level. Similarly, native speakers of English score numerically higher than non-native speakers, although this was not statistically significant. One limitation of this study however, is that there is no indication of group numbers, therefore there could be unbiased groups.

When looking at the accent of the audio, on the other hand, they believed accents of English that are less well known would be harder for learners to understand. In experiment two, the researchers converted two texts into audio recordings of three English accents – U.S. English, Italian and Mandarin Chinese. The participants listened to the audio and then were tested on comprehension. The aim of this experiment was to find out if the accent of the audio has an impact on comprehension. For this experiment, 150 undergraduate students from the University of California in Los Angeles were recruited. The participants were first and second language English speakers. These participants then listened to two of the recordings featuring the same accent counterbalancing using the different texts and audios. Once the participants were split into the groups and had listened to

one text with one audio modality, they had to predict what comprehension score they believed they would receive before partaking in a comprehension test. This process was repeated in the counterbalance. The results all proved non-significant, demonstrating that the accent of the speaker did not have an impact on the listeners' understanding. This was also predicted by the participants themselves as they all predicted their score fairly accurately despite not knowing the questions. Whilst Kiamanesh et al. predicted that the U.S. native accent would be favoured, this was not the case from their results and therefore they concluded that the accent did not significantly impact.

However, although this study shows that text-to-speech can be a useful tool for reading comprehension and does not negatively affect the results when used alongside reading, there are still limitations. One of the main limitations is that it had a between participant design, therefore readers are unsure if the participants themselves had an influence on the results. Additionally, whilst there is a wide sample of participants, with over 100 being present in both studies, they have all been recruited from one undergraduate university and therefore, generalisations may be harder to make due to the participants all seeming to be from one sample pool. Furthermore, this research also makes no indication if the participants had prior knowledge of the texts chosen, as stated by the researchers themselves in their conclusions. Finally, the audio that was shown to the participants was created by the investigator, with no changes in speed or speaker. This may be a limitation due to the implementation of the text-to-speech as most software allows for the change in speed or a tracker of the text to be added depending on what the participant found helpful.

A researcher who found that audio speed is important was Mok (2023). She performed an experiment to investigate which reading style second language learners preferred. In her master's experiment, there were three conditions participants had to complete: reading silently, reading aloud, and listening-whilst-reading. The participants were then asked their views on these conditions and how they felt using them. Five participants were involved in this case-study and were present in the follow-up study as well. For their texts, three texts from previous IELTS (a standard test for English as an additional language) were chosen. All texts were edited to be around 640 words in length and were assessed for a standard language level. The three texts were each shown in one modality to all participants. This means that apparent difference between modalities may be small differences in the text despite attempts being made to prevent this. Questions about the main ideas were asked, with participants answering eight questions in total about each text to gain their understanding, therefore this experiment is an objective measure of performance under each modality. The results of this thesis demonstrated that the views of the application were very individual. Additionally, when looking at their comprehension scores, it showed the wide variety

between each participant. Overall, reading-whilst-listening had the highest mean score and reading aloud the lowest. Participants were also interviewed to gain their opinions on the modalities, and the findings from this also showed it was very individual. The main finding in the interview showed that reading aloud was the most challenging. Across the 5 participants, no consistent pattern was found to say which text modality was the best, and the recording of their experiences showed a similar inconsistent pattern. However, one thing to note was that there were comments made that the audio was too slow (Daisy p.44) or too fast (Ben p.42). Therefore, the audio of the reading-whilst-listening modality played a significant role in the discomfort, with one participant saying they struggled to understand because they "couldn't reread" (Alex p.41). As Lionetti and Cole (2004) state, reading rate and speed of audio can have a great impact on the student's reading and experience. By having the audio at one-speed and preventing the participants from replaying it, Mok (2023) may have created a limitation in her results due to the modality not being used to its full potential, putting pressure on the students and potentially overwhelming them in their study. This qualitative analysis may also explain why, despite the reading-whilst-listening score being the highest, there were mixed views due to the speed. If the speed was correct for each participant, there may have been even more benefit to this modality. However, one reason why some studies have not found substantial evidence of reading-whilst-listening being beneficial could be due to cognitive load. Despite Mayer and Moreno (2003) describing how multiple modes of input could be beneficial, there could also still be an overload of information, especially with the short timeframe between comprehension and testing.

Another thesis that discusses the issues around multimodal learning was written by Lack (2023). In this randomised trial, 27 English language learners participated across two groups to see if asynchronous (participants listen to the audio after reading) or synchronous (participants read and listen at the same time) reading-whilst-listening benefited language acquisition more. One of the hypotheses was that asynchronous would be less taxing on the learner and therefore would allow for more acquisition. The study was designed so the two, randomised groups, could be compared against each other. The participants all had three stories to read which contained 24 target words that had previously been unknown to the participant. A comprehension test and two vocabulary tests that "measured receptive knowledge of orthographic form, and receptive knowledge of form-meaning links" (p.37) were then performed at two points for the data collection, immediately and in a delayed post-test. Their results were surprising because in the immediate post-test, there were no statistically significant differences on the form recognition test, but asynchronous performance was higher on the form-meaning test. In the discussion, they described how this improvement suggests

that asynchronous listening is more effective, and mentioned how this improvement may come due to less of a burden on a learner's cognitive load. As Lack mentions "it is noteworthy that both groups gained much higher scores on both tests than would be expected by chance" (p.57). The fact both groups scored higher is important because it demonstrates that learners can benefit from hearing the text no matter the style. However, one main limitation of the study was the dropout rate between the immediate test and the delayed-post-test. The drop-out rate for the synchronous group was particularly high with 7 and therefore, this prevents any strong conclusions being made with the post-test results. However, despite the drop-out rate, no statistically significant differences were found in the delayed post-test either. In this experimental thesis, the reading-whilst-listening condition is provided via synchronous methods due to this being more commonly used in education and this study not finding any significant difference.

Moussa-Inaty et al. (2011) wanted to see if reading alone had more of a benefit when learning vocabulary as well. In their experiment, native Arabic speakers were asked to learn 38 English words. One control group learnt these words by reading them, whilst the experimental group simultaneously performed reading-whilst-listening. The audio was a recording of a female speaker that was then provided to the participants. The participants had six different activities to perform to try to help their acquisition of these words and then performed two listening tasks to test their knowledge. Testing both groups with listening tests may have led to a limitation for the control group as they only read the text, however, the tests were deemed suitable for the purpose. In this experiment, reading only outperformed the experiment group on both listening questions, which is surprising considering Lack's (2023) study. One reason for this however, as mentioned previously, could be due to the cognitive load and pressure applied with learning vocabulary in a short period of time.

A similar result was found by Hedenström and Barck-Holst (2023). This is the study which is the most relevant to the experiment reported in this thesis. This undergraduate dissertation focused on whether reading-whilst-listening has an impact on English language comprehension in 30 Swedish participants from one university in Stockholm. These participants were all non-native English speakers, however, there was no knowledge of their proficiency level. These 30 participants were split into 3 groups of 10, each assigned to one condition. There was no mention of how allocation took place. The conditions were reading only, reading while listening to a native British speaker, and reading while listening to a text-to-speech recording. The audio was set at a fixed speed for all participants and no additional support was provided, for example text highlighting. It is mentioned in the text that these features were specifically disabled by the investigator. For the recording with

human voice, the audio was recorded previously and provided to the participants alongside the text. The participants completed a reading task in the condition they were allocated on a text chosen from an English exam board paper. They were told they had six minutes to read before answering four comprehension questions. The results from this experiment were unexpected due to there being no significant differences between the conditions for all t-tests performed, going against what the researchers believed to happen as they expected reading-whilst-listening to be more beneficial and for the human voice to be preferred over the text-to-speech software. One reason there may be no evidence of differences could be due to the sample size. By only having 10 participants in each condition, and them only performing the one condition, there may not have been enough participants to prove any hypothesis. Furthermore, alongside the small sample size there is a limitation with participant demographics. All participants were from the same university with similar ages. These are limitations I hope to eliminate in this thesis.

Al-Jarf's (2022) experiment on the benefits of reading-whilst-listening found "significant differences between the experimental and control groups" (p.19). In their 12-week study, 88 students participated during their English language lessons at a university. These participants were grouped based on their classes with 43 students in the experimental group and 45 being present in the control group. At the end of their semester of education, both groups performed two post-tests to see if there were any differences between those who used text-to-speech software alongside their studies and those who did not. Both groups were taught the same material by the same teacher across the 12-week period. Students in the experimental group also performed a qualitative questionnaire to gain their views of the software. Results demonstrated that the experimental group significantly benefited from the text-to-speech addition in the t-tests across the areas tested.

Malone (2024) also found benefits to reading-whilst-listening based on an eye-tracking study to find out how reading is actually performed. In their study, 119 English learners of advanced level were split between a reading only group and reading-whilst-listening. The groups then read the same text which had target vocabulary in. The vocabulary was tested three times in immediate post-tests. Participants were also all tested via eye-tracking across three different areas to see if one modality allowed for better scores. In these results, reading-whilst-listening consistently scored higher in the three vocabulary tests and participants were found to benefit in their reading style too as they seemed to take longer to read new words, but did not re-read as much. One limitation however is that the study was not counterbalanced, meaning the participants only saw the one modality. Furthermore, the audio was controlled by the researcher again, rather than different speeds being available.

2.5 Studies with dyslexic participants in their first language

Enco-Jáuregui et al. (2023) performed a meta-analysis to see whether web accessibility services are beneficial. In their review, they had 22 articles that fit their four research criteria, which they used to answer seven research questions. By having seven research questions and by having vigorous criteria, the researchers gained comprehensive data and were able to back their results with conviction. However, there was bias in their research too, with a main limitation being the temporal criteria. Limiting papers so they had to be published after 2011, meant that there may have been papers discarded that were beneficial. The results of this review showed that all accessibility technologies do have a benefit, with the most prevalent for dyslexics being font size, type and colour. However, as mentioned by Enco- Jáuregui et al. there is a participation limitation in most of the studies they looked at. Due to the nature of the research, gaining a representative sample of dyslexics is quite difficult, and therefore most papers did not manage this.

Additionally, Keelor et al. (2020) performed an investigation to see whether additional accessibility features (text highlighting) benefit dyslexic students when using alongside text-to-speech. Participants read six different texts randomised across the different conditions. This repeated measures design allowed for variability in participants to be minimised, however, the participant burden of all six texts being in one session, with there being a short break time between conditions, may have affected the results for the later conditions. One thing to note is that there were only 10 participants. Therefore, if there was a bigger sample size the results may have been more pronounced and statistically significant. However, the results of this experiment showed that, whilst no condition was better than the others, text-to-speech with highlighting scored slightly better than without. Thus, the current research included text highlighting.

However, the main modality in this thesis is reading-whilst-listening and how text-to-speech software can help with this. Wood et al. (2018), performed an informative meta-analysis looking at how text-to-speech can help dyslexic students comprehend texts. Whilst wood et al. focused on L1 students, the information gained in their study helped form the ideas in the current thesis. In this analysis, 22 studies were included out of 2933 initial records. These 22 remaining studies were used to answer three research questions: What is the average effect size for reading-whilst-listening on comprehension? How much variability is present? How valid and reliable are these studies? The inclusion criteria for this investigation were well structured, with there being 4 different steps for the articles to be incorporated. These steps included a vigorous testing of reading comprehension, the inclusion of an oral section in the experiment, being conducted in a student's second language and

having full-text versions available for the researchers to view. The results from Wood et al.'s meta-analysis demonstrated that the intervention of text-to-speech does lead to an improvement in reading comprehension; however, the effect size of this improvement varies significantly across different studies which led to the conclusion that further investigation was needed. Wood et al. discussed how the main factor in the variation in results was from the different study designs. Whether the experiment was between-subject or between-participants had an influence with between-subject proving to be more effective, thus this design was adopted for the current research. Additionally, they also discussed the quality of previous studies; their findings showed that there is inadequate reporting happening in this area of research with either insufficient statistical reporting or improper controls. Therefore, the main conclusion regarding this meta-analysis is that there needs to be more research into the technology used to help dyslexic students. A recommendation for future research was that speech speed needed to be addressed, and that a larger sample size is needed. In this experiment, participants will be provided different speeds to see if this does play a part in second language knowledge. The recommendation that more research is needed, is still present in more recent literature reviews on this topic. Yaacob et al. (2024) in their brief systematic review also found that there is a benefit to dyslexic students of using text-to-speech software.

Phuc (2020) looked at different reading strategies students who were struggling with comprehension used in a Vietnamese high school, and the difficulties they face. Identifying that General Reading Strategies (GRS) and Specific Reading Strategies (PSRS), have the biggest impact on reading achievement. However, despite the repeated testing using the KMO (Kaiser-Meyer-Olkin) measure and Bartlett's Test where results showed the data is reliable and adequate for testing, there is still a potential for bias in the results with all data being self-reported by the participants.

Draffan et al. (2007), similarly gained self-reported data, performing a telephone study with 455 dyslexic university students. Each student had been provided assistive technology by Student Finance England to help with their studies and the experiment was performed to gain their views on the technology provided. 363 students were provided text-to-speech software (p.109) with 90% of students reporting that they were satisfied with the equipment (p.110). This result is significant because it explains why text-to-speech software is important for learning and that dyslexic students find the equipment beneficial.

2.6 Studies with dyslexic students in their second language

Chai and Chen (2017) performed a systematic review to investigate all the technological assistance provided and used by dyslexics. They found that "text-to-speech technology is the most common"

(p.30), however, their review also showcased that there is a strong bias towards young learners in research. Their articles also focussed on English with most participants having this language as their first language. However, there were only 25 papers investigated in this study with 12 of them mentioning a technology focused on helping reading.

Łockiewicz & Jaskulska (2016) investigated whether text-to-speech software can help with reading and spelling for Polish students in English. 98 participants took part, with 48 having a dyslexia diagnosis. Participants performed three tasks each evaluating a different area of vocabulary; a real word reading task, a nonword reading task and contextualised English words were all examined. This experiment was more focussed on finding out the areas dyslexia students struggle with in their second language, with results showing that "students with dyslexia had more limited L2 vocabulary, regardless of the word difficulty" (p.258) across all three tests performed. This is important because as text-to-speech software has been found to help with vocabulary knowledge and reading in first languages, it can be assumed that it may be the case for second languages too, especially when the difficulties faced are the same.

This assumption was corroborated by Košak-Babuder et al. (2019). In their study they had 47 students with a dyslexia diagnosis, and 233 students without, who were all learning English as an additional language perform three assessments. The students were split between 9 groups in a counter-balanced design due to the groups correlating to the students' year 6 class distribution. The students were all Slovenian and were selected from eight different schools. There were four different texts in this experiment with two being more difficult and two being easier, as established by a text evaluator. The results revealed that the dyslexic students still benefited from the use of text-to-speech software in their additional language.

2.7 Summary

Based on these studies, we can draw multiple conclusions about second language acquisition, reading comprehension, dyslexia and text-to-speech. There has been some research addressing this area of education, however, results have not been conclusive. So far, there has been literature that shows there is no benefit to reading-whilst-listening in additional languages, and there has been research that shows that text-to-speech software has been helpful for reading comprehension and vocabulary understanding. Additionally, it can be said that reading, and the complexity of this skill, is important in second language acquisition, with teachers placing a focus on this skill in classrooms. From research on cognitive load, it is believed that text-to-speech, and using multiple streams of media in general, can be beneficial for learning depending on there not being an overwhelming

amount of unfamiliar content. However, whilst there is evidence for text-to-speech being a helpful technological assistant, there is also evidence to suggest it could put pressure on a student's cognitive load, or that it could create confusion and annoyance for students depending on the speed. Previous research also has limitations in the implementation of reading-whilst-listening software, with the majority of these experiments providing pre-recorded audio set at one speed with no additional features provided. Furthermore, previous research typically did not investigate if any learners have learning disabilities and whether this is a factor that influences data.

2.7.1 Research questions

In light of the literature discussed above, this study aims to investigate English language learners' comprehension when using text-to-speech software and without to see if second language learners can have an improved reading comprehension. To investigate this, we will compare participants' comprehension of a text following reading only or reading-whilst-listening conditions.

Research question 1 (RQ1): Is there a comprehension benefit for English language learners when listening to a text as they read it compared with reading alone?

For this question, the hypothesis was made that reading-whilst-listening should be more beneficial for learners. Therefore, the following hypothesis was created:

H1: Participants will score higher when utilising text-to-speech software than just reading, as measured by immediate post-tests.

Research question 2 (RQ2): Is there a difference between those who could potentially have a learning difficulty (dyslexia) and those without?

For this research question, this hypothesis was made:

H1: Participants with an indication of learning disabilities will benefit more from using text-to-speech software than those without.

Additional research question

Research question 3 (RQ3): Which speed of audio do learners prefer to use when given a choice?

The hypothesis was that beginner learners would benefit more from the slower speed.

3. Methodology

3.1 Study design

In this section, I will provide a comprehensive guide to the research methodology and research design choices that were employed in the creation of this investigation. The present study implemented a repeated measures design with participants randomly allocated to counterbalancing conditions using a randomiser in Gorilla Experiment Builder (Gorilla.sc) (www.gorilla.sc). The data was collected June 2024. The repeated measures approach allowed for all participants to undertake both tasks in both conditions (reading only and reading-whilst-listening). By using this approach, the experiment controls for any accidental differences between the randomly selected participant groups. The counterbalancing also ensures that both texts occur in both conditions. This controls for the situation where one text is easier than the other; having both texts in both conditions ensures the focus is on the condition and not the choice of texts for that condition. Therefore, bias is limited. Random allocation of the participants across the groups was also performed, which has similar benefits to counterbalancing as it makes the groups more equal and helps to limit an overwhelming bias towards one group over the other (Collins et al., 2020; La Caze et al., 2011).

All participants firstly completed a short test designed to measure their language proficiency. The LexTALE test (Lemhöfer & Broersma, 2012), a word recognition test, was chosen for this since previous experiments have shown that the test is an accurate representation of students' language knowledge (Izura et al., 2014; Nakata et al., 2020). Section 2.4.2 explains the LexTALE test in more detail. Following the LexTALE, participants completed two tests to give an indication if they had dyslexia or similar learning difficulties. Section 2.4.3 explains the dyslexia tests in more detail. Participants were then randomly allocated to two groups (Group A and Group B) using the simple randomisation tool in Gorilla.sc (Kang et al., 2008). The groups allowed for the counterbalancing condition as both texts were manipulated across the groups.

Group A saw Text 1 in the reading-whilst-listening manipulation, whilst Group B were shown Text 1 in the reading only condition. The two groups then answered the same six comprehension questions on Text 1. Both groups were then shown Text 2. Group A saw the reading only condition of Text 2 and Group B saw Text 2 in the reading-whilst-listening condition. Groups A and B then had to answer six comprehension questions on Text 2. This data was recorded, and then quantitively analysed using Mann-Whitney U tests and Spearman's Rho tests to answer the research questions.

3.2 Participants

Participants were recruited via two methods. All participants were anonymised in the experiment. Some participants were recruited using a poster (Appendix 8) that was emailed to the schools. After emailing multiple schools and gaining no response, I used some prior connections I had at three schools to gain some participants; these contacts then gave permission to email them the poster for participants. These teachers who received the poster then forwarded it onto the students, allowing them to then decide if they wanted to take part or not, and allowing them to remain anonymous. Participants then emailed to say they wanted to take part, allowing me to send them the link and their unique identification (unique id) so they could take part and their data could be tracked. This approach was used in order to gain access to students from a range of schools, and the diverse pupils in them, allowing for wider data collection and a higher number of differing language backgrounds rather than focusing on a single school.

These participants were unpaid but given the option of taking part in a prize draw for a £100 Amazon voucher for which their personal information was held securely on my University Nexus account and deleted upon completion of the research project.

Since I had not reached the preferred target number of participants through this method, I then also recruited via Prolific. The poster and link to study was uploaded to Prolific (www.prolific.com), where an additional 20 people completed the study. 27 participants had opened and started the experiment, however, 7 did not finish the experiment and therefore were rejected. For this, participants were completely anonymous and were not provided a unique id. Therefore, rather than being included in the prize draw, the participants were paid £9 for their participation.

With the aim of increasing the pool of participants, the demographic for the experiment was not limited by age, first language, or level of English. Due to ethical considerations, participants needed to be over 18, and, according to the CEFR guidelines (Council of Europe, 2014), over B2 level of English, which was established using their LexTALE scores (Lemhöfer & Broersma, 2012). However, these were the sole limitations. The opportunity of recruiting from a wider range of participants allows for more generalisable results compared to prior studies as mentioned in the literature review.

Before the LexTALE test, participants were asked to provide demographic information which included: gender, age in years and months, if they have lived in an English-speaking country, duration of their stay in said country, languages spoken and fluency per language. These questions were all optional. The purpose of gathering this data was to provide a warm-up to using the software and to

provide general information to ensure variety within the pool and that participants fit the recruitment criteria. Of all the participants, 18 participants had reported living in an English-speaking country. See Table 1 for the mean demographics. The participants were asked to self-report throughout these warm-up questions. They were asked to rate their proficiency using the Cambridge version of the CEFR language levels guidelines (www.cambridgeenglish.org, n.d.), on a scale of A1 to C2, through the use of an image included below the question (Appendix 5). The assumption was made that the speakers were honest with their answers, although, according to Olsen et al. (2007), Tanujaya et al. (2022), Lauritsen (1999) and Rosenman et al. (2011), there is a risk of participant bias in self-reporting which could cause data bias. To mitigate this risk, LexTALE was used to check if there is at least a B2 level of word understanding. This test was selected as it was quick to conduct and has previously been shown to correlate with language proficiency. Participants were randomly assigned to the two counterbalancing versions of the study by the Gorilla.sc program.

	Branch A	Branch B	Total
Number of Participants	24	21	45
Mean age	30.01	34.71	32.27
Gender	Male: 8 Female: 14 Prefer not to say: 1 Non-binary: 0	Male: 10 Female: 11 Prefer not to say: 0 Non-binary: 1	Male: 18 Female: 25 Prefer not to say: 1 Non-binary: 1
Self-reported languages	English: 24 Spanish: 9 French: 8 German: 7 Italian: 6 Polish: 5 Portuguese: 5	English: 21 Italian: 13 French: 5 German: 4 Spanish: 3 Greek: 2 Portuguese: 2	English: 45 Italian: 19 French: 13 Spanish: 12 German: 11 Portuguese: 7 Polish:6

	Hungarian: 2 Russian: 2 Arabic: 1 Chinese Mandarin: 1 Japanese: 1 Latin: 1 Latvian: 1 Lithuanian: 1 Swedish: 1	Hindi: 1 Hungarian: 1 Korean: 1 Latin: 1 Polish: 1 Serbian: 1 Turkish: 1 Xhosa: 1	Lithuanian: 1 Hungarian: 3 Greek: 2 Latin:2 Russian: 2 Arabic: 1 Chinese Mandarin: 1 Hindi: 1 Japanese: 1 Korean: 1 Latvian: 1 Serbian: 1 Swedish: 1 Turkish:1 Xhosa: 1
Proficiency level of English	B2: 6 C1: 12 C2: 5	B2: 13 C1: 6 C2: 3	B2: 19 C1: 18 C2: 8
Have lived in English-speaking country	10	6	16
LexTALE score mean	38.83	39.52	39.16
Dyslexia indication results	4	8	12

Table 1: Demographic Questionnaire answers

3.3 Ethics

Before commencing the experiment, participants had to provide informed consent (Appendix 7). This had been approved by the University of Oxford with the approval reference of C1B-24HT-Educ-015 (Appendix 1). Participants were provided this information twice and were asked to confirm their understanding before taking part. The information provided to the participants explained what they were expected to do and outlined the general basis of the tests with reasoning as to why they were asked to participate. They also had the chance to ask any additional questions.

During the data collection in Gorllia.sc itself, participants were all anonymous with them each inputting a unique id given to them for collection of the data. This was clearly laid out to the participants in the information sheet provided. However, as mentioned, the participants had to email me to take part. This information remained on a secure file on my university account and was deleted upon deadline for withdrawal. All that was kept was the email address of the participants. During the data collection via Prolific, all participants remained completely anonymous throughout with no contact or identifiable information being gathered.

3.4 Materials

The experiment was performed on Gorilla.sc. There were five main parts to the experiment - three language tests and two comprehension tests. These were all chosen due to their use in other experiments or their validity in language lessons. This section will explain the experiment and the reasoning behind every element.

3.4.1 Demographic questionnaire

This element focussed on collecting demographic information on the participants which included: gender, age in years and months, if they have lived in an English-speaking country, duration of their stay in said country, languages spoken and fluency per language (Appendix 5). The results of these questions also allowed me to see whether I had completed my research goal of gaining a wider sample group. This demographic questionnaire was created by me, previously tested in my undergraduate dissertation, and trialled to ensure accurate data collection.

3.4.2 LexTALE

After the preview questions, an edited version of the LexTALE test was presented to the participants. The original word recognition task, created by Lemhöfer and Broersma (2012), consisted of participants having 5 seconds allocated to decide whether the 64 terms were existing English

lexemes or non-words. Due to time constraints on this experiment, with having a preference to keeping the study under an hour to lessen burden on participants, some words were omitted and only 49 words were included in the script. This was deemed as an adequate number for language level understanding to be gathered due to the comprehension level of language used (Albrechtsen et al., 2008; Bachman & Palmer, 1996; Daller & Phelan, 2007; Nation, 2006; Stæhr, 2008). The LexTALE was chosen due to its reoccurring use in experiments and although there are limitations, the language level of the learner was not the main part of the experiment and therefore it was deemed that this test is sufficient for the task selected.

3.4.3 Dyslexia tests

After the LexTALE, participants performed two different dyslexia tasks. These were chosen as they had previously been used in similar experiments within the literature and demonstrated when participants could have dyslexia. Following analysis, it was clear that combining the two would yield more accurate results without participants having to disclose private information. The first test was an English sentence verification task that was created and used by Garvin and Krishnan (2021) to test L1 English. Due to this, the test has not been verified or established for L2 use, therefore I decided to also include the questionnaire for increased validity.

In the verification test, participants were shown 50 sentences which were either true or false. In three seconds, they had to read the phrase and click the correct answer. There were three trial participants: one with dyslexia and two without. Their results were in line with Garvin and Krishnan. In fact, the participant with dyslexia, showed lower performance than the others. One participant without dyslexia was L2 English, and they still scored higher than the one with dyslexia. This shows the validity of Garvin and Krishnan in this scenario. Consequently, when discussing this section of the experiment with the participant with dyslexia, she commented on the speed at which the sentences were shown, asking for additional processing time.

The second test was a short questionnaire (*Dyslexia Test*, n.d.) in which the participants could rate some potential difficulties dyslexic students tend to have. The ten questions were provided to the participants along with a scale, similar to a Likert scale, for them to choose where they fell in these situations. See Appendix 9 for the questions.

3.4.4 Headphone check

A headphone check with Huggins' Pitch and browser sound check was then performed using a template created by Milne et al. (2021). All participants passed this check which ensured that they

would hear the audio in the relevant text condition. This test has been used in other research (Cramer & Huggins, 1958; Stärk et al., 2023; Woods et al., 2017) and therefore was deemed a reliable check for participants.

3.4.5 Texts 1 and 2

For this experiment, two texts and associated questions were chosen from worksheets by different companies for use in schools. Both were checked to see if they were at the correct level by choosing texts from workbooks tagged at B2 level. Both fictional texts told a story and had six following questions which followed the same style (4 grammatical and 2-word choice) and were both suitable for the language learners needed for this study. Neither text assumed the reader to have any prior knowledge on the subject.

Text A (Appendix 2) was taken from LearnEnglish British council (LearnEnglish, 2020). This text was edited slightly because the text was too long for the task, however, the short, deleted sections of the texts were not essential in preserving the central narrative of the story. This text was focused on an unsuccessful astrological experiment and its repercussions, which was chosen as it is a topic not taught in classrooms. This text was longer and made up of 868 words: the counterbalancing means that the discrepancies in difficulty between the texts are controlled.

Text B (Appendix 3) was taken from English Download workbook (*English Download [B2]*, n.d.). This text had one additional paragraph added to it, to make the word count 651. Whilst the book is set for B2 level learners, when researching the words and knowledge needed, I believed this was more C1 level. After considering the vocabulary and tenses covered in the textbook compared to those in the text, it was deduced this was more C1 level; and therefore, I felt a slightly lower word count was more beneficial due to the complexity and the counterbalancing.

Both texts had six similar questions where the first four were multiple choice about the content and grammar present in the stories (Appendix 2.1 and 3.1). The last two questions involved the participants choosing the correct word to fill in the sentence, making sure they used the correct grammatical tense as well. All the information needed to complete these tasks was found in the text, therefore it did not test the student's grammar as the sentence to fill in was taken directly from the text. These questions were verified using other textbooks, such as English file (*English File 4th Edition Pre-Intermediate Students Book - Studocu*, n.d.), as a common style of comprehension questions the learners would have likely seen during their own lessons. Therefore, these questions taken from the workbooks and adapted slightly to be similar in style, were deemed valid comprehension questions.

3.4.5.1 Video and audio

For this experiment, I had to find a way to include the audio for the reading-whilst-listening manipulation. Ideally, I would provide specialist dyslexic software to all participants to allow for the most reliable test. However, due to the funds available for this project, this was not a viable option. Therefore, I decided to create videos I could integrate into Gorilla.sc which simulated the experience of using software like this (Appendix 4). These videos were created by making recordings of me using a personal copy of ClaroRead Plus (*ClaroRead - Text to Speech Software | Claro Software*, n.d.). This software read out the texts with a screen ruler highlighting the line of text that is being read. This software is commonly provided as support for dyslexic students in the United Kingdom and therefore was deemed valid for this experiment. The audio speaker chosen was called Hollie, and the description is of a middle class, middle aged, British female. This software was beneficial for the different aspects I wished to simulate for this experiment, for example, different audio speeds. The reader could choose the speed of the audio; the audio is a good replication of spontaneous human conversations and there is a screen ruler which is present on the screen that can highlight each line of text as it is being read. To simulate the reader having a choice of speed, I made three recordings for each text at the three speeds and once the participants reached this point in the experiment, participants were able to choose which speed they wanted to watch the video at.

In the software, the screen ruler can be changed to different colours, however, due to the software having to be implemented using a video created by me and not being provided to all participants, I had to choose one colour for this experiment. Research by Jakovljevi´c et al. (2021) shows that yellow is one beneficial screen addition colours for dyslexics, and research by Niklaus et al. (2023) demonstrates the use and benefit of screen rulers. Therefore, I decided to use the yellow colour for the screen ruler in this experiment.

Finally, I want to acknowledge that there are limitations to implementing the reading-whilst-listening manipulation in this way. The main limitation is that the participants are unable to make changes to the audio themselves: the colour of the screen ruler and voice were chosen by me, and whilst I did try to implement different speeds for the participants, when using the software themselves, the participants would be able to manipulate the speed along a wider range of speeds. I will return to discuss the implications of these differences in the Discussion (section 5).

3.5 Procedure

3.5.1 Flow diagram

Figure 1: Flow diagram

As shown in this flow diagram, one difference between all the groups is the speed of audio that the participants choose. This was added due to Mok's (2023) experiment where her results found a discrepancy between whether the participants liked the speed, found it too slow, or found it too fast. Therefore, for my experiment participants were able to choose whether they wanted to hear the audio at -0.25, at 0 – common British speech, or at +0.25, according to ClaroRead Plus's measurements. The version of ClaroRead used was 8.0.7. The participant could then choose which speed they believed was more beneficial for them in this task.

Each section of this experiment had introductions beforehand explaining what to expect and what was expected of them. Although the provided rules may not have been adhered to, and I was unable to verify, every attempt was made to make sure they were followed with numerous mentions of the rules and expectations. Additionally, to make the test as fair as possible, participants were asked to read the text, and watch the video of the text, just once without taking notes. This was done to demonstrate a more controlled reading style and to limit the time the participants took on the tests. The instructions shown outlined what the participants should expect in the next section and what answers they could provide.

3.5.2 Cognitive pre-testing trials

For the trials, three volunteers, including one who has been diagnosed with dyslexia, all participated at three different stages of the experiment building to gain their feedback. During the trials, all three participants used different devices, and none found any specific issues due to the device they were completing the experiment on by the final iteration. Therefore, it was believed that this experiment could be performed on phone, tablet and laptop.

The results from the first stage granted the biggest change to the experiment. After the trial, one of them asked whether just the audio as an automatic start and end with the text on screen was a valid representation of how ClaroRead Plus is usually used by dyslexics. As I agreed with her question, my supervisor and I discussed this concept, and we concluded to upload a full video with a screen ruler being present on the screen too.

After editing the experiment to add in the video, the three volunteers completed the experiment again. All three found a great improvement with the screen ruler with two of them making a comment that they found it easier to keep track of where they were in the text. A few modifications were pointed out with some spelling mistakes and bolding of text to be made. There was also found to be a technical problem with Text A's video when the speed was at -0.25 as the continue button was not present and therefore, they could not advance in the experiment originally.

After making those corrections, this final trial confirmed that the experiment was suitable for its use. They had no corrections after completing this trial and therefore it was deemed ready for the main experiment. All participants said they found no technical difficulties and finished within the allocated 60 minutes, and that the experiment was easy to perform and understand. All three participants stated that they had enough time to read the texts, and that the videos were adequate for their purpose. The participants also liked the different choices of speed and said they worked well for allowing them to choose the best one for them.

3.6 Main experiment

Once the trials had all been completed and ethical clearance granted, I started the main experiment. For this, I emailed different second language schools around England (Appendix 10) and gained responses from four teachers who were willing to forward the poster (Appendix 8) and the information sheet (Appendix 6) to their students, three of whom were previous contacts from teaching experience. 35 participants then emailed me regarding the experiment, and I sent them the link along with their unique id as previously explained. Of these students 25 were included. Due to the small sample size this granted, I then uploaded the link of the experiment onto Prolific to gain the remaining 20 who were included.

3.7 Data analysis

Once the data was gathered, analyses were conducted using the IBM SPSS Statistics software (Version 29.0.1.0). My dependent variable was the average performance in the comprehension tests, and my independent factors were reading only or reading-whilst-listening. I also performed a within participant test to check the effects of reading, with or without the audio. I measured how well overall participants did between groups within texts with the different conditions.

The variable of audio speed was only compared against each other in choice, the focus was on the comprehension differences and the speed variable was added to allow for participants to choose the speed they felt more comfortable with so the comprehension discrepancies could be from the audio variable and not due to participants not having a speed they felt was a positive addition rather than a hinderance.

I performed Mann-Whitney U tests and Spearman's Rho tests on the data to compare whether there was a statistical improvement. These tests were chosen because the results followed what was necessary for these tests to be valid. The results will be discussed in Section 4.

4. Results

In this section, I will present the results of this experiment, explaining that the addition of reading-whilst-listening positively impacted Text 1 score and did not influence Text 2 comprehension scores. The results are structured around the three research questions. Research question 1 was the primary question and had a clear directional prediction, whilst the other two questions were exploratory in nature. All statistical tests were performed using SPSS (version 29.0.1.0) and are written up and performed following Field (2017).

4.1 Demographic results

The dataset initially had 59 participants but 14 were rejected due to not completing the full experiment. Of the remaining 45 participants, 18 were male, 25 were female, one was non-binary and one preferred not to say. The mean age of these participants was 32.27 years, who all spoke English at B2 level or above. 12 participants scored over 500 on the dyslexia indication test and were therefore categorised as dyslexic for the purpose of research question 2. The two groups were split across two branches to allow for a counterbalanced study. The two groups were randomised by the Gorilla.sc program itself. In Branch A there were 24 participants and in Branch B there were 21 participants who completed the experiment and whose results were thus included.

4.2 RQ1: Is there a comprehension benefit for English language learners when listening to a text as they read it compared with reading alone?

4.2.1 Summary of mean data in both groups:

The surprising finding in this research was the results for Text 2. These statistics demonstrate that both texts are non-significant. However, Text 1 audio had a positive impact on the comprehension with there being an increase of 1.46 in the mean correct scores when audio was added. In Text 2, audio did not facilitate understanding and instead there was a 0.81 decrease in the mean score of participants. The questions for this test were out of 6, meaning the highest possible score was 6/6.

Text 1 no audio mean score	3.29
Text 1 audio mean score	4.75
Text 2 no audio mean score	5
Text 2 audio mean score	4.19

Table 2: summary of mean score across all factors.

4.2.2 Descriptive statistics

In this section, I will describe the statistics created for the texts, which led to the decision of which tests to perform. The benefit of these statistics is the clear and concise summary of the data provided.

The first descriptive statistics created were for Text 1, the space text. Statistics were created based on the scores the two groups gained out of six – the maximum score they could get. Text 1 Reading only had a mean score of 3.29 with a standard deviation of 1.62, as shown in Table 2. Text 1 Reading and audio, on the other hand, had a mean score of 4.75 and a standard deviation of 1.45. Figure 2 shows this increase quite clearly with even the error bars not overlapping, meaning that the audio significantly impacted this text. Additionally, the descriptive statistics demonstrated that the data for reading with audio was negatively skewed (skewness -1.85) with a kurtosis of 4.14. Since the data were not normally distributed the tests reported below were non-parametric tests. The data for reading only was also negatively skewed (skewness of -0.44) but the kurtosis was -0.13, meaning that it was evenly distributed. Consequently, since some of the data was not normally distributed, for consistency the tests performed were non-parametric tests throughout.

Text 2, on the other hand, led to some surprising statistics. This text was about our voices and was determined to be more difficult than Text 1 due to the language used. Descriptive statistics were created again based on the score gained out of six by the participants of each condition. Text 2 reading only had a mean score of 5 with standard deviation being 1.474. The unexpected result was that Text 2 reading-whilst-listening scored 4.19 as their mean. This is lower than expected and goes against the hypothesis of this investigation. The standard deviation for this text was 1.569, so whilst this means that the two conditions overlap on their standard deviation error bars, as shown in Figure 2, the chance of error still does not indicate an improvement in comprehension when provided audio. Furthermore, Text 2 had different descriptive statistics than what was found with Text 1 which meant the tests performed could have all been parametric tests. As mentioned however, all tests

were non-parametric to keep the results fair. Reading only had a skewness score of -1.42 and a kurtosis score of 1.18, meaning all the data was evenly distributed. Text 2 reading-whilst-listening had a skewness score of -0.35 and a kurtosis score of -0.99.

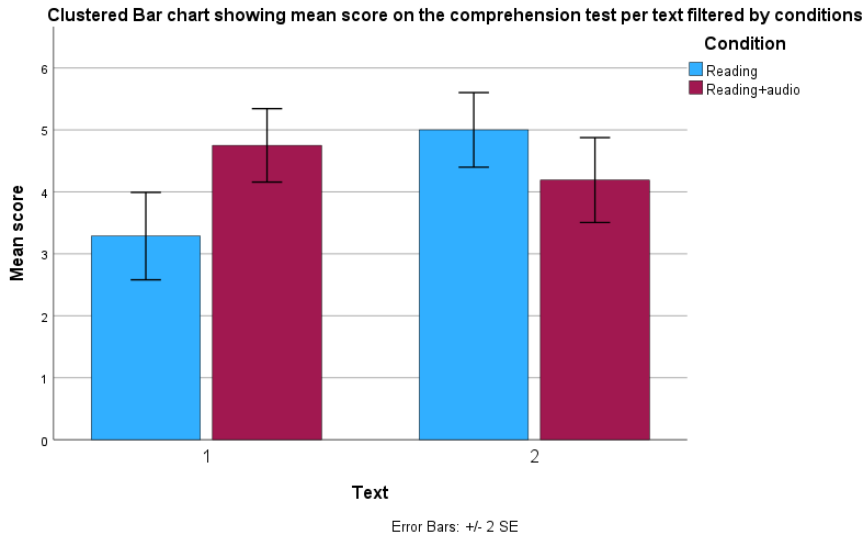


Figure 2: Plot of means.

4.2.3 Condition comparisons

As shown in Table 2, the conditions affected the comprehension differently depending on the text. Both texts had 6 questions the groups could score full marks on, meaning the maximum score was 6 and the minimum 0. Numerically, Text 1 saw an increase in performance and Text 2 saw a decrease when in the condition with audio. The differences in these scores cannot be due to differences in the participant groups due to the counterbalancing of my study and the random allocation between groups. Participants who saw Text 1 in the condition format saw Text 2 in the reading-whilst-listening condition, therefore the group influence was small as both groups performed better in one condition. Whilst there may still be some limitations in the groups, as will be discussed in section 6.3, the counterbalancing allows for all participants to try both conditions and both texts. Furthermore, the use of the error bars in Figure 2 demonstrates the range of scores the participants gained and therefore allows for more in-depth analysis based on what can be seen, showing that audio on Text 2 was still less beneficial than Text 1.

In this section, I will explain the approach taken to test the differences between the audio and reading only conditions, to see whether the results led to any significant data. I will be examining both texts individually to see if the different conditions allow for the difference in results. Since each

text had both reading only and reading-whilst-listening conditions performed by different groups, a non-parametric test was performed to compare the means of the two independent samples. Despite predicting a benefit of audio, due to the descriptive statistics a two-tailed test was performed on both texts.

4.2.3.1 Text 1

An independent-samples Mann-Whitney U test was performed to compare the text with and without audio against each other due to the negative skewness. The test indicated that there is a statistically significant difference between the comprehension scores of the participants for Text 1 without audio (Mdn = 3.29) and with audio (Mdn = 4.75), U = 114.5, z = -3.204, p = 0.001. Consequently, the null hypothesis of audio not helping, is able to be rejected with the assumption being made that the results would not have occurred by chance. The conclusion of this test is that the use of audio in this text did have a significant impact for language learners.

4.2.3.2 Text 2

Due to the even distribution of data for Text 2, t-tests were able to be performed. The test indicated that there is no significant difference between the comprehension scores of the participants for Text 2 without audio (Mdn = 5) and with audio (Mdn = 4.19), t(43) = 1.783, p = .0.082. Thus, the results were not significant, and I cannot reject the null hypothesis.

However, due to Text 1 having non-normally distributed data, a non-parametric test was also performed on Text 2 for consistency. A Mann-Whitney U test was conducted. The results of this test were also non-significant demonstrating that the addition of audio when reading did not facilitate learning in Text 2, U = 171, z = -1.94, p = 0.052

4.3 RQ2: Is there a difference between those who could potentially have a learning difficulty (dyslexia) and those without?

In this section, I will first look at the means for the groups of students who meet the threshold of scoring above 500 in the two-part dyslexia test. This was scored based on the sum of their scores on the sentence verification and their self-provided scores for the questionnaire. The demographical data for this group was smaller, with only 12 participants who scored over 500. Of these 12 participants, 4 were in Group 1 and 8 were in Group 2, consequently, the samples are considered too small to warrant performing statistical tests on the separate groups.

For this experiment, I first split the data to show those with dyslexia and those without to obtain the descriptive statistics. The mean scores for this sample group were surprising yet again, however, as they did not follow the predicted pattern. As shown in Figure 3, the mean comprehension score for Text 1 when the participants had indicators of dyslexia was 3 when reading only and 5.25 when reading-whilst-listening whilst the Text 2 mean was 4.88 when reading only and 4.14 when reading-whilst-listening. The error bars present on this figure demonstrate that for Text 1, audio still would have helped for these participants, whilst the error bars on Text 2 overlap more. Contrastingly, Figure 4 shows the mean comprehension scores when participants did not have indicators of dyslexia. In Text 1, the mean was 3.37 when reading only and 4.65 when reading-whilst listening, whilst Text 2 was 5.06 when reading only and 4.21 when reading-whilst-listening.

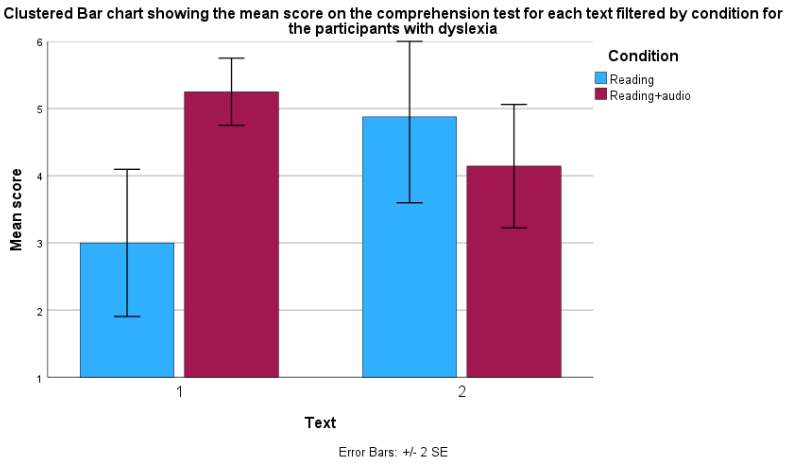


Figure 3: Plot of means for those with dyslexia.

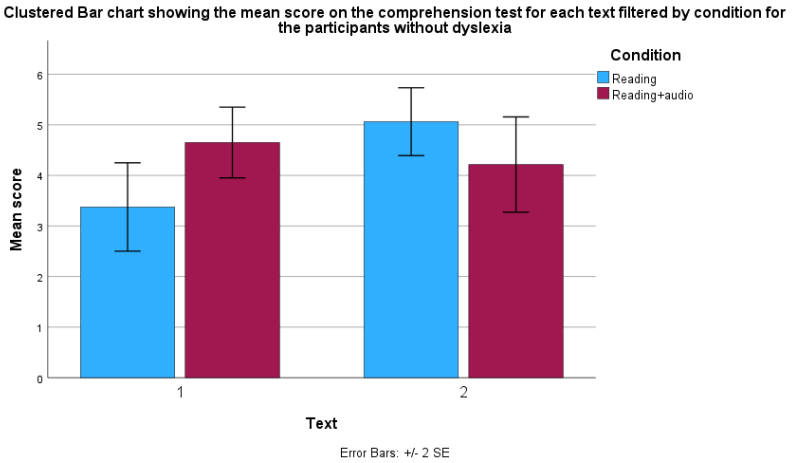


Figure 4: Plot of means for those without dyslexia.

To answer the RQ, based on the mean score values, a Spearman's rho was performed to see if there was any evidence that the benefit of audio is greater in participants with an indication of dyslexia. This test was performed across all participants due to there being no clear prediction as to which participants would benefit more from reading-whilst-listening. To look at the relationship for each participant the difference, between their score in the reading plus audio condition was computed compared to their score in the reading only condition. These scores were then correlated against the indication score. If the participant scored highly on the indication score, they were deemed to be more likely to have dyslexia. Therefore, a positive correlation would be seen in the direction of participants benefitting more from the addition of audio. This was not found in this test r(45) = [-0.097], p = [0.527]. As shown in Figure 5, this does not lead to a relation being present between the participants' score on the dyslexia test and their difference in score when audio is added.

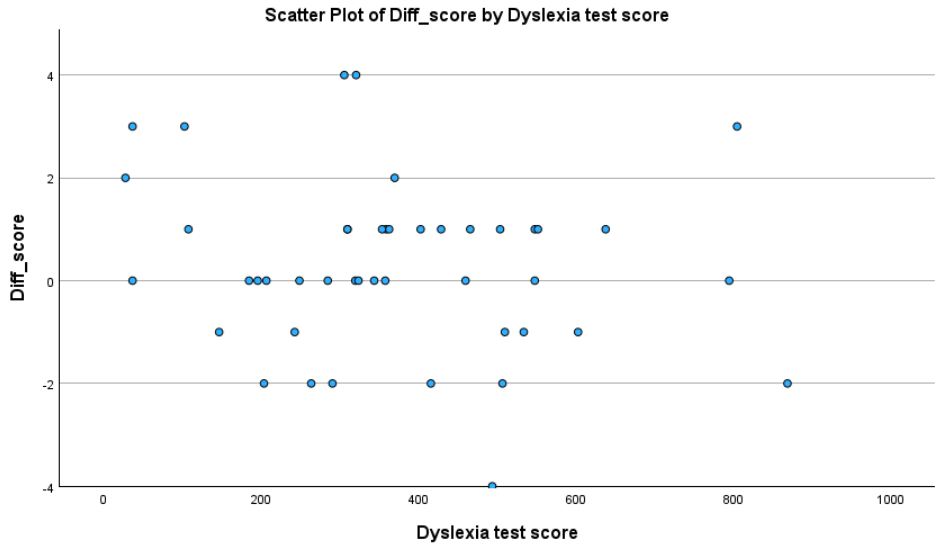


Figure 5: Scatter plot of the participants score on the dyslexia test and their difference in score when audio is added.

4.4 RQ3: Is there a relationship between speed choice and comprehension scores?

My final RQ investigated if there was any significance between the different speed choices of the participants. This question focussed on which speed was chosen by each participant, to see if providing participants with the choice made an impact. Once the data was gathered, I was curious to see if the different speeds led to a different score on the audio tests. Participants who read Text 2 with audio all chose the default speed "0", according to ClaroRead Plus's speed choices this is equal to common conversation speed. Text 1 with audio, however, had more of a split in their data with

participants choosing all three available options. Due to the small number of participants in this section, and uneven spread across the speeds, it was deemed inappropriate to conduct statistical tests. However, as shown by Table 3, the means are in the direction of a faster speed leading to better scores, but the difference appears small. This could be an area for future research, as will be discussed in the conclusion (section 6).

4.4.1 The results of the speed choices for Branch A

The descriptive statistics of those who chose the speed for Text 1 were distributed between all three choices, but the group sizes were not equal. The mean score for all three groups was consistent, showing that the significant difference is too small to comment on.

Speed	Number of participants	Mean score on text 1
-25	2	3
0	12	3.83
+25	10	4

Table 3: descriptive statistics of speed choices in Text 1

5. Discussion

Introduction

This section will discuss and critically evaluate the findings from the statistical tests performed. The results will be related and compared to those of previous experiments discussed in the literature review and will be focused on the research questions. This section is focused on each of the research questions and how the findings relate to what has previously been reported. Ultimately, this section demonstrates the relevance of the results and the impact this research has in the wider educational sector.

RQ1

The first research question focused on the comprehension of the participants and whether the use of text-to-speech to facilitate reading-whilst-listening helped or hindered the comprehension. This concept was tested by the use of immediate reading comprehension tests. Analysis of these scores and the mean results suggests mixed results. The findings for Text 1 showed a positive growth in comprehension when an audio recording of the text was listened to alongside reading, whilst Text 2 showed the opposite. Initial observations of Text 1 were that they were consistent with the prediction that audio is helpful. However, Text 2 revealed contradictory results for the hypothesis. The change in comprehension score for Text 1 was +1.46 whilst Text 2 was –0.81. The most natural interpretation is accidental differences between participant groups which I will examine more in this section.

A statistical test was performed on the mean scores, with a Mann Whitney U test resulting in Text 1 having a significant result of 0.001; an assumption can be made that the results would not be found by chance. This result is consistent with the finding that text to speech software has been found to be useful in students with dyslexia (Handley 2009). More broadly, this means that the assumption that it can help more than just the intended societal group is correct. A consistent finding in other areas that tools which benefit first language comprehension for individuals with disabilities can also be beneficial to addition language comprehension (Liao et al., 2020). Alternatively, previous studies investigating the benefits of additional audio on reading comprehension such as Hedenström and Barck-Holst (2023) have not consistently found evidence of this benefit. Consequently, a discussion into what accounted for the difference took place. One idea was the different participant demographics. In this experiment, there was a larger sample of participants who all came from a

wider demographic of previously known languages. Another idea could be from the texts and questions used. In this experiment, the materials chosen were tailored for classroom use, whilst previous experiments typically used materials for examinations (Kiamanesh et al., 2023). Furthermore, research where a negative impact has been found also leads to questions (Drezek, 2001). In Drezek's (2001) experiment, the results were not significant, but were repetitive with the experiment group scoring lower every time. In the current experiment, the participants were only tested once and the scores compared, therefore, repeated testing may directly influence the results.

Text 2, contrastingly, had the opposite results which where that the audio addition decreased reading comprehension. The results of the Mann Whitney U were $p = 0.052$. This is not in accordance with the findings of Al-Jarf (2022) whose experiment with 30 participants found significant gains with the use of text-to-speech software. However, this result is in relation to other studies, for example Drezek (2001). The result here showing that audio had a negative impact on the comprehension for this text was surprising as it was assumed that audio would be beneficial and that overall scores would be higher.

Therefore, given these findings, the most natural interpretation is that we are not seeing an effect of audio helping additional language learners, but one group of participants is naturally stronger than the other as the same group does best regardless of the audio. This mixed finding is in relation to previous literature and means there is still no decisive result as to whether reading-whilst-listening helps second language learners' reading comprehension or not. The key concept that will now be discussed is the possible reason why the predicted pattern was not seen. As with any study, there are two possibilities when a null effect is found. One is that there truly is an effect (H1 is true) but an effect was not seen because of some aspect of the way the study was done. The other possibility is that there truly is no effect and adding audio is actually not specifically helpful to English language learners. These two statements will now be evaluated.

Starting with the first possibility, while every attempt was made to make the test as fair as possible, and the participants were divided into groups randomly, there may have accidentally been some differences between the groups. This could be due to the different demographics of the participants. With the demographic questionnaire, some information was collected to show participants' language knowledge, time spent in an English-speaking country and even in the English language level of the participants. However, upon exploration of the demographical information gathered, the

results did not reveal any obvious differences between the groups apart from language proficiency which will be considered next. Therefore, the absence of participants' languages showing any clear answers, suggests that there may be other factors influencing the results which I may not have gathered. This is therefore an interesting idea for future investigation. Furthermore, the other known languages of participants may influence their listening ability and therefore influence their scores. However, due to the results in this investigation, more demographic information could be collected in the future. The idea of language knowledge influencing understanding was investigated by Kiamanesh et al. (2023), although they found that whilst native speakers performed better, their two non-native groups performed similarly in how much audio affected them.

Another way the groups may have influenced the scores is from the proficiency level. This is the most obvious explanation of the results; however, it could be a limitation of my randomisation implementation. As shown in Figure 2 in the methodology chapter, the participants in Branch 1, who saw Text 1 in the reading only condition and Text 2 in the reading-whilst-listening condition, had a higher language proficiency distribution and therefore showed an increase with audio being added. Whilst the participants were randomly allocated, Branch 1 gained more participants at C1 level with there being 12 at C1 compared to 6 in Branch 2. Consequently, despite the statistical tests being implemented for the texts rather than the branch, there is still evidence of this group benefiting from the addition of audio more than the other branch. Therefore, there is a potential limitation in how the groups were split leading to this split influencing the scores. Collins et al. (2020) and La Caze et al. (2011) both explain the benefits of randomisation and the fact any randomisation is good despite the risk of accidental differences, particularly in smaller sample studies. An alternative for future studies could be to use stratified randomisation.

Another aspect could be due to the tests themselves. All participants completed the same 12 tests with the participants being shown the questions immediately after the texts. However, an unexpected finding when looking at the data was that question 5 for Text 2 caused the most incorrect answers for both groups. Whilst this may not have influenced the whole results, it demonstrates that the texts may have caused the difference in results if there was misunderstanding. The question was, "Of course, people rarely want to change their voices ----, even if it was possible." The four answer choices to fill this gap were adapted from the workbook, but the answer was written in the text, and they had two similar options for the word choices for the participants to choose between: "completely" which was the correct answer, and "slight" which was incorrect both grammatically and with the text however was the 2nd most chosen answer. Whilst

this was overall the hardest question, and therefore it was not the sole reason this group performed worse than the other, it is still something worth noting. This is especially obvious as, with closer focus on the group's answers, Text 2 proved harder for the group with audio overall compared to when Text 1 had audio.

Another reason that there was no effect and there was an influence on the answers when reading-whilst-listening for Text 2 could be because of cognitive overload. This may not be the most likely possibility; however, it is still an interesting point to consider. The questions themselves were not read out with audio to the participants like the texts were; however, previous literature by Kim (2020) found that participants preferred the addition of text-to-speech for easier texts rather than more difficult ones. As mentioned in the methodology section 2.4.5, Text 2 was deemed to be slightly more challenging than Text 1 due to a consideration of the vocabulary and tenses covered in the text compared to those in a B2 level workbook. This text was therefore deduced to be more a low C1 level or very high B2 and therefore, is the more challenging text out of the two. Kim's (2020) findings are therefore replicated in this experiment with audio being more helpful for the easier text.

It is possible that the way in which this study implemented the simulation of ClaroRead was ineffective, and that if we had been able to give each student a copy of the actual software we would have found it to be beneficial for comprehension. Whilst, in this experiment, the software was implemented in the most natural way with the constraints, the rigid nature of the audio by it being prerecorded and unable to be implemented in person, could have led to these results. In this experiment, ClaroRead software was added to the text by the investigator. This was then recorded three times and the recording of the text was uploaded into the experiment. Therefore, the participants were unable to utilise the software in the most beneficial way for them. Whilst changes were made compared to previous studies with the addition of speed and the text highlighting, every participant is individual and therefore may have different needs that were not met. This is what Mok (2023) found with her experiment. Of her 5 participants, 3 mentioned that the reading-whilst-listening condition was not the right speed for them and therefore it created more difficulties rather than aiding comprehension. In the current study, we aimed to avoid this difficulty by allowing participants a choice of audio speed. However, this may still not have been sufficiently flexible and participants may still have not had audio that was at an appropriate speed for them. Additionally, the use of the screen ruler, or even the choice of using Hollie, a British female audio voice, may have influenced the enjoyment and usefulness in this study.

Finally, an acknowledgment must be made that null results do not mean we know the null is true. Despite 45 participants from a range of backgrounds taking part, the experiment still may not have had enough power to show the anticipated results and if there were more participants, or a smaller demographic sample that focusses on one area of population in a larger number, a positive effect may be established. It might be that if there had been a large enough sample, an audio benefit in both texts may have been seen. As shown by Amer (1997), having a larger sample size of students being read to by the teacher compared to other studies, led to more comprehension in the students.

The second possibility is that additional audio during reading, while helpful for first language dyslexic students, is not in fact beneficial for additional language learners. If this is the case, it would be an extremely interesting finding when there is research to show reading-whilst-listening is beneficial for first language speakers. However, it could still be the reason for the results found in this experiment. One reason could be due to second language learners typically finding listening harder than reading (Melby-Lervåg & Lervåg, 2014). Rintaningrum (2018) explained this in their study when they looked at 40 student surveys to find out what students exactly find difficult about listening to English. Verhoeven (1990) also demonstrated that listening is challenging due to the age difference of learning; children learn to read once they already know some language, whereas we learn additional languages at the same time as learning to speak and therefore the text needs to be at the same level as the language knowledge.

Another reason on this same area is on comprehension itself and how this could be due to additional language learners traditionally finding listening comprehension harder than reading comprehension. Rubin (1994) examined this and explained how comprehension needs to be engaging for participants and learners so they can benefit more from the skill. Despite the audio being used alongside reading in this study, the addition of the audio may have hindered participants and made them focus more on what they were hearing, resulting in a struggle between the skills being tested. The idea that listening comprehension being harder for second language learners is something that also could be further researched.

Cognitive overload could be another factor at play here. Whilst previous research in second language learning has found that reading-whilst-listening reduces cognitive load and that having multiple aspects of input can assist learners (Mayer & Moreno, 2023), there is also evidence that reading-

whilst-listening can hinder cognitive load due to the dual-input. Moussa-Inaty et al. (2011) also found this in their second language study where reading alone lead to the most gains. Whilst the addition of audio can be useful for learners, each learner is individual and therefore the audio may need to be tailored for them to make sure they are gaining the most.

Furthermore, as found by Malone (2024), when reading silently/aloud participants can go at their own speed and check vocabulary, whereas, when using audio to facilitate reading comprehension you must go at the speed of the audio, and it is more challenging to pause and look up unknown words. Therefore, when presented with unknown words in these texts, the addition of audio may not have facilitated understanding and instead created more hinderances. In this case, adding audio would not be helpful for learning and comprehension due to the terminology being unable to be comprehended anyway. This is an important point in relation to the differences between first and second language learning. In a student's first language, it is typically assumed they know most of the words they are reading in their spoken form, but in additional languages, when dealing with novel words they might not know the spoken form and therefore the addition of audio would not help.

RQ2

Research question 2 investigated whether there is a difference between those who could potentially have a learning difficulty (dyslexia) and those without. The results of this experiment were the most surprising because of the contrast to previous research. The mean scores for this test showed a large increase for Text 1 when audio was added (Mdn went from 3 to 5.35) however, Text 2 showed a decrease again (Mdn went from 4.88 to 4.14). Due to the number of participants who scored high enough on the indication tests being so small, and the interesting mean scores, a Spearman's Rho was performed on all participants to try to see any correlations. The results of this correlation test showed no correlation between the dyslexia indication total score and the difference score between the text that had audio and the text without for each participant $r(45) = [-0.097]$, $p = [0.527]$. Whilst the texts were in-line with the previous research question and those participants with dyslexia benefitted for Text 1 and not for Text 2, this result did not align with prior research. The results of this question may be due to the small sample size who scored high enough on the two dyslexia probability tests, however, this may have been due to the tests being performed in English, as will be discussed more in the limitation section.

As previous research (Chai & Chen, 2017; Wood et al., 2018) has often found, dyslexic learners in both their first and second languages often benefit from reading-whilst-listening. The use of text-to-

speech is common and promoted and leads to comprehension gains for this group of learners. Therefore, the surprising results in this study may be due to the sample size and the implementation of software.

RQ3

This research question looking into which speed participants chose to use for the reading-whilst-listening modality, is very specific to the participants due to fact the allocation to groups in this section is not random. Participants chose which speed group they wanted to be a part of and so the groups were created specifically, and results cannot be generalized. For Text 2, all participants chose to use the default speed: speed 0 according to ClaroRead Plus' speed settings. Text 1 did have more variation with at least one participant choosing each speed option, however the majority still chose to use speed 0. No statistical tests were able to be completed due to the split of participants in each group being small, however, the sum of participants choices were calculated and findings showed that more participants chose the faster speeds. The mean score was higher the faster the speed although score difference was minimal. One possibility for this result may be that all participants have chosen the best speed for them, as they were meant to.

Due to comments made in previous literature, this question was included and investigated. Both Lionetti and Cole (2004) and Mok (2023) have shown that the audio speed is important when reading-whilst-listening. Mok (2023) had 3 participants mention their discomfort with the audio speed which hindered their learning. Lionetti and Cole (2004) did prove that the different audio speeds did not influence the scores, however, they did influence participants' comfort. Consequently, in this experiment, I wanted to allow participants to have more volition in the experiment, in comparison to these examples, allowing participants to choose which speed gave them the most confidence.

Conclusion

When considering the relationship between the results of this experiment, the findings of previous research and the research questions themselves, it is safe to say that no assumptions can be made. There are so many mixed results, even in experiments themselves.

Similarly, when thinking of the themes throughout this dissertation, it appears that the expectation of reading-whilst-listening leading to an increase in reading comprehension is something that possibly needs further exploratory investigation to find a strong conclusion on whether reading-

whilst-listening directly benefits second language speakers or not. Benefits of this study included the wider demographic and the choice of audio speed, however, further research is needed and a balance between the language level of the speaker and the input needs to be the same. This conclusion all depends on each individual learner and their own choices and requirements, therefore, the variations in person need to also be studied and accounted for more closely.

6. Conclusion

In this dissertation, I have investigated whether there is an immediate comprehension benefit for reading-whilst-listening using text-to-speech software. The three research questions helped differentiate the different areas of the study and below will be a short evaluation of the success in answering these questions. Data was gathered using an online experiment builder, Gorilla.sc. The data gathered from this two-part counterbalanced experiment was analysed using SPSS to evaluate the influence, scope and value of this research. This section will summarise the findings and evaluate the experiment including reference to limitations and will provide avenues for future research before examining the pedagogical implications.

6.1 The research summarised

This investigation was led by two main hypotheses and one additional question. These hypotheses are restated below, with the key findings of the results summarised.

RQ1: Is there a comprehension benefit for English language learners when listening to a text as they read it compared with reading alone?

The results of this section show inconclusive data on whether reading-whilst-listening helps second language comprehension or not. The mean scores for Text 1 show an increase and the mean scores for Text 2 show a decrease when audio is added. Therefore, it cannot be confidently stated whether it improves learning.

RQ2: Is there a difference between those who could potentially have a learning difficulty (dyslexia) and those without?

The results of this hypothesis were hard to address because there is no discernible difference between those who were likely to have dyslexia and those without. Of the participants, only a small sample fit within the criteria and therefore there was not enough data to fully test. Therefore, no evidence of reading-whilst-listening benefitting learners with a learning disability was established. However, another reason could be that the participants answered the questions in their additional language they were learning and so, the benefit could be for everyone rather than just the select few. In this experiment, a correlation was not found between those who potentially have dyslexia and their score when audio is added.

RQ3: Which audio speed do learners prefer to use when given a choice?

This question was an additional factor into this experiment due to comments made by participants in previous experiments (see literature review section 3). It was believed that participants would prefer having multiple speed options depending on which they found most helpful. The surprising result was that most participants chose audio speed 0 – the equivalent to normal speech, with audio speed +25 being the second most popular. All participants in Branch 2, where Text B had the audio, chose audio speed 0 and therefore all statistical tests were only performed on Branch 1. The means revealed no major correlations.

6.2 Pedagogical implications

From this research, some preliminary pedagogical implications can be drawn due to the result from RQ1 where it was found that there is an added benefit for the comprehension text scores of language learners using text-to-speech software alongside reading.

The results suggest that learners should be encouraged to read whilst listening to text as the additional input allows for a more detailed comprehension. The software used in this experiment could be implemented in classroom computers which allow students to use the software during English language instructed lessons. Additionally, a recording of the text using the software could also be shared with the students, allowing them to utilise the benefits without having to download anything.

6.3 Limitations and future research

Finally, there are limitations to this experiment, and these will now be discussed with implications for future research.

Firstly, participants had a limited audio speed choice compared to if they actually had the software. Therefore, whilst there was more choice granted to the participants than in previous studies, allowing them to decide which they believed would be most helpful and potentially limiting the problems of the audio speed found in the past, the issue was more rigid than it would be in a real-life situation.

Furthermore, the way the software was implemented meant that all three ClaroRead Plus additions were all used and implemented by the researcher: the audio sound, the addition of the text-to-

speech and the coloured line. Due to financial constraints, the participants were not each granted access to ClaroRead Plus, and therefore could not implement the software themselves, meaning that the additional changes that are allowed to be tailored to each person were not present, and so the software was a limitation in itself. This is not a completely unbiased data set subsequently, due to the restriction of the recording of the text using the software. The impact this may have on the dataset is that some participants may still have had problems with the audio, the colour of the overlay may have created more distractions or the addition of the audio itself may have had technical issues. Whilst this was deemed to be an acceptable way to implement the software for this experiment, future research may provide each participant with their own software, allowing them to make their own choices on the voice, speed and volume and allowing them to follow the text with the ruler more closely. The main point of this idea is that if participants had been provided their own software and been able to tailor the application themselves, then I believe their comprehension would have been better and there would have been even greater differences between the conditions. Therefore, the small effect size, despite it being a larger study than previously performed, may reflect the limitations of my implementation of ClaroRead software.

Additionally, this research is believed to be the first wide-scale experiment with participants from multiple demographics. Therefore, more research is needed to verify the results. However, despite the diversity in the participants, there was a discrepancy of language and cultural representation. Some groups had 1 participant who spoke a language, whilst others, like Italian, had 19. Future research should aim to replicate the findings with different participants and a more equal sample spread. This research also had more participants with a higher language proficiency in one group, despite attempts at randomisation. Therefore, future research could investigate whether the unbalanced language proficiency in groups led to the results.

Whilst this experiment was focused on second language proficiency, testing the participants in English as well as having the texts in English may have influenced the results as well. This limitation may be small, however, it could have provided additional challenges for the learners. Consequently, future research could tailor the experiment so the tests are in the participants first language.

Another area of research should focus more on the addition of text-to-speech over a longer period of time. In this experiment, time constraints meant that the comprehension of the texts could only be tested the once, immediately after participants read the text. It would be interesting to see if a

longitudinal study would identify whether regular, continuous use of reading-whilst-listening skills leads to benefits in learning and retention of the vocabulary and knowledge.

Furthermore, whilst every attempt was made to prevent differences, with a counterbalanced study where the participants were randomly allocated between groups and all participants saw both conditions in different texts, there may still be group bias present in the results. Demographic distribution and language knowledge may influence. Similarly, whilst counterbalancing is deemed an effective way for data to be controlled in this experiment, it may not eliminate all external factors. Participants may still be fatigued, bored or the effect of order may still influence results. Therefore, an idea for future research is to repeat the experiment with different participants with breaks, different orders and pre-testing of their knowledge.

Finally, whilst it has been shown that reading-whilst-listening can have a benefit for these participants when used in their own time and used with headphones, it would be interesting to see if there is still a benefit in a more controlled classroom setting. Due to socio-economic factors of pupils and learners around the world, there may be technological divides preventing some students from using this software when not in classrooms. Therefore, future research could investigate other ways this technology could be implemented to see if the benefit is present in other situations, especially with the cost factor of providing every student with text-to-speech software.

References

Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). Lexical knowledge, lexical inferencing and writing. In *Vocabulary and writing in a first and second language: Processes and development* (pp. 160–194).

Albright, L. K., & Ariail, M. (2005). Tapping the Potential of Teacher Read-Alouds in Middle Schools. *Journal of Adolescent & Adult Literacy*, *48*(7), 582–591. https://doi.org/10.1598/JAAL.48.7.4

Al-Jarf, R. (2022). Text-To-Speech Software for Promoting EFL Freshman Students' Decoding Skills and Pronunciation Accuracy. *Journal of Computer Science and Technology Studies*, *4*(2), Article 2. https://doi.org/10.32996/jcsts.2022.4.2.4

Amer, A. A. (1997). The effect of the teacher's reading aloud on the reading comprehension of EFL students. *ELT Journal*, *51*(1), 43–47. https://doi.org/10.1093/elt/51.1.43

An analysis of Japan's English as medium of instruction initiatives within higher education: The gap between meso-level policy and micro-level practice | Higher Education. (n.d.). Retrieved 2 April 2024, from https://link-springer-com.ezproxy-prd.bodleian.ox.ac.uk/article/10.1007/s10734-018-0323-5

Aoki, N. B., Cohn, M., & Zellou, G. (2022). The clear speech intelligibility benefit for text-to-speech voices: Effects of speaking style and visual guise. *JASA Express Letters*, *2*(4), 045204. https://doi.org/10.1121/10.0010274

Asrimawati, I., & Margana, M. (2020). Using reading while listening in English as a foreign language (EFL) learning context. *JEES (Journal of English Educators Society)*, *5*. https://doi.org/10.21070/jees.v5i2.840

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford University Press. https://contentstore-cla-co-uk.ezproxy-prd.bodleian.ox.ac.uk/EReader/Index?guid=0c21116f-8048-ed11-a27c-

0050f2f06092&pcid=3084857&t=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJqdGkiOiIwMmRmM2RlMi0wODA1LTQ0N2MtYjAzZC00Nzg1YTc5NDllYTEiLCJuYmYiOjE3MjIyNjAyNjEsImV4cCI6MTcyMjI2MDU2MSwiaWF0IjoxNzIyMjYwMjYxLCJpc3MiOiJEQ1MiLCJhdWQiOiJSZWFkZXIifQ.TVEGjqWhBBAHIT9TWeq8G9sh9T04nWH74I_HeeAm0hk&b=False

Bajrami, L., & Ismaili, M. (2016). The Role of Video Materials in EFL Classrooms. *Procedia - Social and Behavioral Sciences*, *232*, 502–506. https://doi.org/10.1016/j.sbspro.2016.10.068

Bamford, J., & Day, R. R. (1998). *Extensive reading in the second language classroom*. Cambridge University Press.

Bassetti, B. (2008). Orthographic input and second language phonology. In *Input matters in SLA* (pp. 191–206).

Beck, I. L., & McKeown, M. G. (1991). Research Directions: Social Studies Texts Are Hard to Understand: Mediating Some of the Difficulties. *Language Arts*, *68*(Nonfiction, Language Learning, and Language Teaching), 482–490. https://doi.org/10.58680/la199125316

Bhola, N. (2022). Effect of Text-to-speech Software on Academic Achievement of Students with Dyslexia. *Integrated Journal for Research in Arts and Humanities*, *2*(4), Article 4. https://doi.org/10.55544/ijrah.2.4.45

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching Subtitled Films Can Help Learning Foreign Languages | PLOS ONE. *PloS One*, *11*(6). https://journals.plos.org/plosone/article/?id=10.1371/journal.pone.0158409

BORRÁS, I., & LAFAYETTE, R. (2011). Effects of Multimedia Courseware Subtitling on the Speaking Performance of College Students of French. *The Modern Language Journal*, *78*, 61–75. https://doi.org/10.1111/j.1540-4781.1994.tb02015.x

Busse, V., & Walter, C. (2013). Foreign Language Learning Motivation in Higher Education: A Longitudinal Study of Motivational Changes and Their Causes -. *The Modern Language Journal*, *97*(2), 435–456. https://doi.org/0.1111/j.1540-4781.2013.12004.x

Carrell, P. L., Devine, J., & Eskey, D. E. (Eds.). (1988). *Interactive Approaches to Second Language Reading*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524513

Chai, J. T., & Chen, C. (2017). A Research Review: How Technology Helps to Improve the Learning Process of Learners with Dyslexia. *Journal of Cognitive Sciences and Human Development*, *2*. https://doi.org/10.33736/jcshd.510.2017

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

Chang, A. C.S. (2011). The Effect of Reading While Listening to Audiobooks: Listening Fluency and Vocabulary Gain. *Asian Journal of English Language Teaching*, *21*, 43–64.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264–75278. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2988510

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

Cho, K.-S., & Hey-Jung, K. (2004). Recreational Reading in English As a Foreign Language in Korea: Positive Effects of a Sixteen-Week Program—ProQuest. *Knowledge Quest*, *32*(4), 36–38.

ClaroRead—Text to Speech Software | Claro Software. (n.d.). Texthelp. Retrieved 20 May 2024, from https://www.texthelp.com/en-gb/solutions/dsa/claroread/

Collins, R., Bowman, L., Landray, M., & Peto, R. (2020). The Magic of Randomization versus the Myth of Real-World Evidence. *New England Journal of Medicine*, *382*(7), 674–678. https://doi.org/10.1056/NEJMsb1901642

Council of Europe. (2014). *The CEFR Levels*. Common European Framework of Reference for Languages (CEFR). https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Cramer, E. M., & Huggins, W. H. (1958). Creation of Pitch through Binaural Interaction. *The Journal of the Acoustical Society of America*, *30*(5), 413–417. https://doi.org/10.1121/1.1909628

Dai, L., Kritskaia, V., Van Der Velden, E., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). Evaluating the usage of Text-To-Speech in K12 education. *Proceedings of the 2022 6th International Conference on Education and E-Learning*, 182–188. https://doi.org/10.1145/3578837.3578864

Daller, H., & Phelan, D. (2007). What is in a teacher's mind? Teacher ratings of EFL essays and different aspects of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 234–244). Cambridge University Press. https://doi.org/10.1017/CBO9780511667268.016

Daskalovska, N. (2014). Incidental Vocabulary Learning through Reading. *Journal of Linguistics and Language Teaching*, *5*(1), Article 1.

Dawson, K., Antonenko, P., Lane, H., & Zhu, J. (2019). Assistive Technologies to Support Students With Dyslexia. *TEACHING Exceptional Children*, *51*(3), 226–239. https://doi.org/10.1177/0040059918794027

Day, R., Bamford, J., Renandya, W., Jacobs, G., & Yu, V. (1998). Extensive Reading in the Second Language Classroom. *Relc Journal*, *29*, 187–191. https://doi.org/10.1177/003368829802900211#

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. Instructional Science, [online] 38(2), pp.105–134. doi:https://doi.org/10.1007/s11251-009-9110-0.

Discovering Statistics Using IBM SPSS by Andy Field—Google Search. (n.d.). Retrieved 29 July 2024, from https://www.google.com/search?q=Discovering+Statistics+Using+IBM+SPSS+by+Andy+Field&rlz=1C1CHBF_en-GBGB917GB919&sourceid=chrome&ie=UTF-8

Dousay, T. A. (2016). Effects of redundancy and modality on the situational interest of adult learners in multimedia learning. Educational Technology Research and Development, 64(6), 1251–1271. https://doi.org/10.1007/s11423-016-9456-3

Draffan, E. A., Evans, D. G., & Blenkhorn, P. (2007). Use of assistive technology by students with dyslexia in post-secondary education. Disability and Rehabilitation. Assistive Technology, 2(2), 105–116. https://doi.org/10.1080/17483100601178492

Drezek, J. (2001). *Adult ESOL reading comprehension and text-to-speech software* [M.A., The University of Texas at Arlington]. Retrieved 1 December 2023, from https://www.proquest.com/docview/304708248/abstract/98DD08E5646B4DE9PQ/1

Duncan, K. (2022). *Assistive Technologies for Students with Dyslexia*. https://pressbooks.pub/techcurr20221/chapter/assistive-technologies-for-students-with-dyslexia/

Dyslexia Test. (n.d.). Dyslexia | Dyslexic Advantage. Retrieved 20 May 2024, from https://www.dyslexicadvantage.org/dyslexia-test/

Enco-Jáuregui, L., Meneses-Claudio, B., & Auccacusi-Kañahuire, M. (2023). Web accessibility for people with dyslexia: A systematic literature review. *EAI Endorsed Transactions on Pervasive Health and Technology*, *9*, 1–12. https://doi.org/10.4108/eetpht.9.4274

English Download [B2]: Workbook. (n.d.). Retrieved 20 May 2024, from https://online.pubhtml5.com/rohc/yfca/#p=1

English File 4th edition Pre Intermediate Students Book—Studocu. (n.d.). Retrieved 20 May 2024, from https://www.studocu.com/en-gb/document/university-of-oxford/english/english-file-4th-edition-pre-intermediate-students-book/52664394

Field, A. (2017). *Discovering Statistics Using SPSS* (Fifth edititon). SAGE Publications Ltd.

Garvin, B., & Krishnan, S. (2021). Curiosity-driven learning in adults with and without dyslexia. *Quarterly Journal of Experimental Psychology*, *75*(1). https://doi.org/10.1177/17470218211037474

Gorilla Experiment Builder. (n.d.). Gorilla Experiment Builder - Easily Create Online Behavioural Experiments. [online] Available at: https://www.gorilla.sc/.

Gotesman, E., & Goldfus, C. (2010). The Impact of Assistive Technologies on the Reading Outcomes of College Students with Disabilities. *Educational Technology*, *50*.

Govender, A., & King, S. (2018). Using Pupillometry to Measure the Cognitive Load of Synthetic Speech. *Interspeech 2018*, 2838–2842. https://doi.org/10.21437/Interspeech.2018-1174

Grunér, S., Östberg, P., & Hedenius, M. (2018). The Compensatory Effect of Text-to-Speech Technology on Reading Comprehension and Reading Rate in Swedish Schoolchildren With Reading Disability: The Moderating Effect of Inattention and Hyperactivity Symptoms Differs by Grade Groups. *Journal of Special Education Technology*, *33*(2), 98–110. https://doi.org/10.1177/0162643417742898

Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, *51*(10), 906–919. https://doi.org/10.1016/j.specom.2008.12.004

Harvey, J., Hux, K., Scott, N., & Snell, J. (2013). Text-to-speech technology effects on reading rate and comprehension by adults with traumatic brain injury. *Brain Injury*, *27*(12), 1388–1394. https://doi.org/10.3109/02699052.2013.823649

Hatami, S., & Tavakoli, M. (2012). The Role of Depth versus Breadth of Vocabulary Knowledge in Success and Ease in L2 Lexical Inferencing. *TESL Canada Journal*, 1–1. https://doi.org/10.18806/tesl.v30i1.1123

Hedenström, E., & Barck-Holst, A. (2023). *Investigating reading comprehension in Reading While Listening and the relevancy of The Voice Effect*. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-333662

Hempenstall, K. (1997). The Role of Phonemic Awareness in Beginning Reading: A Review. *Behaviour Change*, *14*, 201–214. https://doi.org/10.1017/S0813483900003314

Hoffman, J. V., Roser, N. L., & Battle, J. (1993). Reading Aloud in Classrooms: From the Modal toward a 'Model' on JSTOR. *The Reading Teacher*, *46*(6), 496–503.

IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *PSICOLOGICA*, *35*(1), Article 1.

Jakovljevi´c, T., Jankovi´c, M. M., Savi´c, Andrej M., Soldatovi´c, I., Coli´c, G., Jakulin, T. J., Papa, G., & Kovi´c, V. (2021). The Relation between Physiological Parameters and Colour Modifications in Text Background and Overlay during Reading in Children with and without Dyslexia. *Brain Sciences*, *11*, 539–558.

Kalaitzi, A. (2020). *Reading comprehension in the second language. The relationship between second language reading comprehension, second language vocabulary and first language reading comprehension* [Master thesis]. https://www.duo.uio.no/handle/10852/80432

Kang, M., Ragan, B. G., & Park, J.-H. (2008). Issues in Outcomes Research: An Overview of Randomization Techniques for Clinical Trials. *Journal of Athletic Training*, *43*(2), 215–221. https://doi.org/10.4085/1062-6050-43.2.215

Keelor, J., Creaghead, N., Silbert, N., Breit-Smith, A., & Horowitz-Kraus, T. (2018). Language, Reading, and Executive Function Measures as Predictors of Comprehension Using Text-to-Speech.

Reading & Writing Quarterly, *34*(5), 436–450. https://doi.org/10.1080/10573569.2018.1486764

Keelor, J. L., Creaghead, N. A., Silbert, N. H., Breit, A. D., & Horowitz-Kraus, T. (2023). Impact of text-to-speech features on the reading comprehension of children with reading and language difficulties. *Annals of Dyslexia*, *73*(3), 469–486. https://doi.org/10.1007/s11881-023-00281-9

Keelor, J. L., Creaghead, N., Silbert, N., & Horowitz-Kraus, T. (2020). *Text-to-Speech Technology: Enhancing Reading Comprehension for Students with Reading Difficulty*. *14*.

Kiamanesh, L., Murphy, D., & Castel, A. (2023). To Read or to Listen: The Effect of Text-to-Speech Software and Accents on Comprehension of Digital Media. *Modern Psycological Studies*, *29*(1), 1–29.

Kim, J., & Shin, W. (2014). How to Do Random Allocation (Randomization). *Clinics in Orthopedic Surgery*, *6*(1), 103–109. https://doi.org/10.4055/cios.2014.6.1.103

Kim, P. (2020). The Effects of LWR (Listening While Reading) and Intrinsic Cognitive Load on L2 Reading Comprehension. *Journal of Second Language Acquisition and Teaching*, *26*.

Koda, K. (2005). *Insights into Second Language Reading: A Cross-Linguistic Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524841

Koda, K. (2009). Learning to Read in New Writing Systems. In *The Handbook of Language Teaching* (pp. 463–485). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444315783.ch25

Koh, J. (2024). Deconstructing the benefits of reading-while-listening on L2 reading comprehension: The influence of cross-orthographic distance. *Foreign Language Annals*, *57*(1), 163–183. https://doi.org/10.1111/flan.12732

Košak-Babuder, M., Kormos, J., Ratajczak, M., & Pižorn, K. (2019). The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners. *Language Testing*, *36*(1), 51–75. https://doi.org/10.1177/0265532218756946

Krashen, S. (1985). Chapter 1: The Imput Hypothesis. In *The Imput Hypothesis: Issues and Implications* (pp. 77–109). https://www.uio.no/studier/emner/hf/iln/LING4140/h08/The%20Input%20Hypothesis.pdf

Krashen, S. (2013). *Second language Acquisition: Theory, applications, and some conjectures*. Cambridge University Press.

Ktistakis, E., Gleni, A., Tsilimbaris, M. K., & Plainis, S. (2024). Comparing silent reading performance for single sentences and paragraphs: An eye movement-based analysis. *Clinical and Experimental Optometry*, *107*(4), 449–456. https://doi.org/10.1080/08164622.2023.2237974

La Caze, A., Djulbegovic, B., & Senn, S. (2011). What does randomisation achieve? *Evidence-Based Medicine*, *17*, 1–2. https://doi.org/10.1136/ebm.2011.100061

Lack, C. (2023). *Incidental Vocabulary Acquisition from Synchronous Reading-while-Listening vs. Asynchronous Reading-and-Listening: A Randomised Trial*. [Master's Thesis] University of Oxford.

Langland-Hassan, P. (2021). Inner speech. *WIREs Cognitive Science*, *12*(2), e1544. https://doi.org/10.1002/wcs.1544

Lauritsen, J. L. (1999). Limitations in the Use of Longitudinal Self-Report Data: A Comment Exchange. *Criminology*, *37*(3), 687–694.

LearnEnglish. (2020). *First star I see tonight*. https://learnenglish.britishcouncil.org/general-english/story-zone/b2-c1-stories/first-star-i-see-tonight-b2/c1.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *National Library of Medicine*, *44*(2), 325–343.

Liao, S., Kruger, J. L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *Journal of Specialised Translation*, *33*, 70–98.

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, *21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Lionetti, T. M., & Cole, C. L. (2004). A Comparison of the Effects of Two Rates of Listening While Reading on Oral Reading Fluency and Reading Comprehension. *Education and Treatment of Children*, *27*(2), 114–129.

Łockiewicz, M., & Jaskulska, M. (2016). Difficulties of Polish students with dyslexia in reading and spelling in English as L2. *Learning and Individual Differences*, *51*, 256–264. https://doi.org/10.1016/j.lindif.2016.08.037

Lu, X. (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*, *96*(2), 190–208.

MacKay, D. G. (1992). Constraints on Theories of Inner Speech. In *Auditory Imagery*. Psychology Press.

Malone, E. (2022). Primary foreign languages national curriculum expectations in England: Implications for practice from the Ofsted curriculum research review for languages and the 24 languages subject inspections. *Education 3-13*, *0*(0), 1–12. https://doi.org/10.1080/03004279.2022.2106287

Malone, J. E. (2024). Toward a Theory-Based Account of the L2 Vocabulary Processing and Learning Benefits of Reading While Listening [Ph.D., University of Maryland, College Park]. In *ProQuest Dissertations and Theses*. https://www-proquest-com.ezproxy-prd.bodleian.ox.ac.uk/docview/3074097508/abstract/14A93C000AD642E5PQ/1

Mayer, R. E. (2002). *Multimedia learning—ScienceDirect*. *41*, 25–139.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Melby-Lervåg, M. and Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. Psychological Bulletin, 140(2), pp.409–433. https://doi.org/10.1037/a0033890.

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Mohamed, A. A. (2018). EXPOSURE FREQUENCY IN L2 READING: AN EYE-MOVEMENT PERSPECTIVE OF INCIDENTAL VOCABULARY LEARNING. *Studies in Second Language Acquisition*, *40*(2), 269–293. https://doi.org/10.1017/S0272263117000092

Mok, E. (2023). *Inner and outer voices: L2 readers' experiences of silent reading, reading aloud and reading-while-listening* [Master's Thesis] University of Oxford.

Moon, D. (2012). Web-Based Text-to-Speech Technologies in Foreign Language Learning: Opportunities and Challenges. In T. Kim, J. Ma, W. Fang, Y. Zhang, & A. Cuzzocrea (Eds.), *Computer Applications for Database, Education, and Ubiquitous Computing* (pp. 120–125). Springer. https://doi.org/10.1007/978-3-642-35603-2_19

Moussa-Inaty, J., Ayres, P., & Sweller, J. (2011). 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.

Moussalli, S. (2022). *Exploring Intelligent Personal Assistants in Second Language Acquisition* [Phd, Concordia University]. https://spectrum.library.concordia.ca/id/eprint/990606/

Nakata, T., Tamura, Y., & Aubrey, S. (2020). Examining the Validity of the LexTALE Test for Japanese College Students. *The Journal of Asia TEFL*, *17*(2), 335–348. https://doi.org/10.18823/asiatefl.2020.17.2.2.335

Nasri, M., & Biria, R. (2017). Integrating Multiple and Focused Strategies for Improving Reading Comprehension and L2 Lexical Development of Iranian Intermediate EFL Learners.

International Journal of Applied Linguistics and English Literature, *6*(1), 311–321. https://doi.org/10.7575/aiac.ijalel.v.6n.1p.311

Nation, I. (2006). How Large a Vocabulary is Needed For Reading and Listening? *The Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary* (1st ed). Heinle, Cengage Learning.

Nation, I. S. P. (2005). Teaching and Learning Vocabulary. In Handbook of Research in Second Language Teaching and Learning. Routledge.

Niklaus, A. G., Cai, T., Bylinskii, Z., & Wallace, S. (2023). Digital Reading Rulers: Evaluating Inclusively Designed Rulers for Readers With Dyslexia and Without. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, *47*, 1–17.

Nooteboom, S. (2010). Monitoring for speech errors has different functions in inner and overt speech. In M. B. H. Everaert, Ø. Nilsen, H. N. M. De Mulder, & T. Lentz, *The Linguistics Enterprise* (pp. 213–233). John Benjamins Publishing Company. https://www.torrossa.com/en/resources/an/5015300

Oktalia, D., & Drajati, N. (2018). English teachers' perceptions of text to speech software and Google site in an EFL Classroom: What English teachers really think and know. *International Journal of Education and Development Using ICT*, *14*(3). https://www.learntechlib.org/p/188288/

Olsen, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, *18*(4), 287–291.

Park, C. (2009). Implementing Computer-Assisted Language Learning in the EFL Classroom: Teachers' Perceptions and Perspectives. *International Journal of Pedagogies and Learning*, *5*. https://doi.org/10.5172/ijpl.5.2.80

Paribakht, T.S., & Wesche, M. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. Canadian Modern Language Review, 53(1), pp.13–40. https://doi.org/10.3138/cmlr.53.1.13.

Parr, M. (2012). The Future of Text-to-Speech Technology: How Long before it's Just One More Thing we do When Teaching Reading? *Procedia - Social and Behavioral Sciences*, *69*, 1420–1429. https://doi.org/10.1016/j.sbspro.2012.12.081

Phuc, N. (2020). A Study On Reading Strategies Used By Vietnamese High School English Language Learners. *International Journal of Scientific and Research Publications (IJSRP)*, *10*, 255–263. https://doi.org/10.29322/IJSRP.10.06.2020.p10231

Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive Load Theory*. Cambridge University Press.

Prolific (2023). Prolific · Quickly find research participants you can trust. [online] www.prolific.com. Available at: https://www.prolific.com/.

Rintaningrum, R. (2018). Investigating reasons why listening in English is difficult: voice from foreign. *Asian EFL Journal*, *20*(11), 6-15.

Roberts, K., Takahashi, K., Park, H.-J., & Stodden, R. (2013). Text-to-Speech Use to Improve Reading of High School Struggling Readers. *The International Journal of Literacies*, *19*, 89–97. https://www.researchgate.net/profile/Leni-Marlina-4/publication/287047127_Learning_English_as_Foreign_Language_in_Indonesia_through_English_Children's_Literature/links/59ffd257aca272347a2af55f/Learning-English-as-Foreign-Language-in-Indonesia-through-English-Childrens-Literature.pdf#page=104

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*, *2*(4), 320–332. https://doi.org/10.1504/IJBHR.2011.043414

Rubin, J. (1994). A Review of Second Language Listening Comprehension Research. The Modern Language Journal, 78(2), p.199. https://doi.org/10.2307/329010.

Ruben, R. J. (1997). A Time Frame of Critical/Sensitive Periods of Language Development. *Acta Oto-Laryngologica*, *117*(2), 202–205. https://doi.org/10.3109/00016489709117769

Rupley, W. H., Nichols, W. D., Rasinski, T. V., & Paige, D. (2020). Fluency: Deep Roots in Reading Instruction. *Education Sciences*, *10*(6), Article 6. https://doi.org/10.3390/educsci10060155

Sakai, H., & Kikuchi, K. (2009). An analysis of demotivators in the EFL classroom. *System*, *37*(1), 57–69. https://doi.org/10.1016/j.system.2008.09.005

set.et-foundation.co.uk. (n.d.). The importance of cognitive load theory | Society for Education and Training. [online] Available at: https://set.et-foundation.co.uk/resources/the-importance-of-cognitive-load-theory#:~:text=Cognitive%20Load%20Theory%20%28CLT%29%20-%20coined%20in%201988.

Sohler, S. (2020). *Developing Listening Comprehension in ESL Students at the Intermediate Level by Reading Transcripts While Listening: A Cognitive Load Perspective* [M.A., Brigham Young University]. https://www-proquest-com.ezproxy-prd.bodleian.ox.ac.uk/docview/2441550705/abstract/A5A1209A7124481APQ/1

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*(2), 139–152. https://doi.org/10.1080/09571730802389975

Stärk, K., Kidd, E., & Frost, R. L. A. (2023). Close Encounters of the Word Kind: Attested Distributional Information Boosts Statistical Learning. *Language Learning*, *73*(2), 341–373.

Stein, D. S., Blum, N. J., & Barbaresi, W. J. (2011). Developmental and Behavioral Disorders Through the Life Span. Pediatrics, 128(2), 364–373. https://doi.org/10.1542/peds.2011-0266

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257–285.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem-solving in mathematics and science. Journal of Educational Psychology, 81, 457–466.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4, 295–312.

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), Cambridge handbook of multimedia learning (pp. 159–167). Cambridge, UK: Cambridge University Press.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. Cognition and Instruction, 12, 185–233.

Syrdal, A. K., Bunnell, H. T., Hertz, S. R., Mishra, T., Spiegel, M., Bickley, C., Rekart, D., & Makashay, M. J. (2012). Text-to-speech intelligibility across speech rates. *Interspeech 2012*, 623–626. https://doi.org/10.21437/Interspeech.2012-194

Tanujaya, B., Prahmana, R. C. I., & Mumu, J. (2022). Likert scale in social sciences research: Problems and difficulties. *FWU Journal of Social Sciences*, *16*(4), 89–101.

The Construct of Multilingualism in Language Testing. (n.d.). Retrieved 27 December 2023, from https://www.tandfonline.com/doi/epdf/10.1080/15434303.2019.1680679?needAccess=true

Torgerson, D. J., & Torgerson, C. J. (2008). Systematic Reviews of Randomised Controlled Trials. In D. J. Torgerson & C. J. Torgerson (Eds.), *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction* (pp. 160–174). Palgrave Macmillan UK. https://doi.org/10.1057/9780230583993_16

Turker, S., Fumagalli, B., Kuhnke, P., & Hartwigsen, G. (2024, June 19). The 'reading' brain: Meta-analytic insight into functional activation during reading. https://doi.org/10.31219/osf.io/4bu8p.

Vanderplank, R. (2016). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, *49*(2), 235–250. https://doi.org/10.1017/S0261444813000207

Verhoeven, L. T. (1990). Acquisition of Reading in a Second Language. *Reading Research Quarterly*, *25*(2), 90–114. https://doi.org/10.2307/747596

Verlaan, W., & Ortlieb, E. (2012). Reading While Listening: Improving Struggling Adolescent Readers' Comprehension Through the Use of Digital Audio Recordings. *What's Hot in Literacy Yearbook*, 30–37. https://www.texasreaders.org/uploads/8/6/6/5/8665759/literacy_summit_yearbalook_2012_-_final.pdf#page=38

Wai Wai, L., Kee Jiar, Y., & Handayani, L. (2023). A systematic review on interventions for children with dyslexia. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(3), 1674. https://doi.org/10.11591/ijere.v12i3.25099

Waring, R., & Nation, P. (2004). SECOND LANGUAGE READING AND INCIDENTAL VOCABULARY LEARNING. *ANGLES ON THE ENGLISH-SPEAKING WORLD*, *4*, 11–23.

Wesche, M., & Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *The Canadian Modern Language Review*, *53*(1), 13–40. https://doi.org/10.3138/cmlr.53.1.13

What are the different 'levels' of learning a language? (n.d.). Www.Cambridgeenglish.Org. https://www.cambridgeenglish.org/learning-english/parents-and-children/how-to-support-your-child/what-are-the-different-levels-of-learning-a-language/.

What dyslexics see. (n.d.). Dyslexia the Gift. Retrieved 5 May 2021, from https://www.dyslexia.com/question/what-dyslexics-see/

Wissinger, D. R., Truckenmiller, A. J., Konek, A. E., & Ciullo, S. (2024). The Validity of Two Tests of Silent Reading Fluency: A Meta-Analytic Review. *Reading & Writing Quarterly*, *40*(2), 118–134. https://doi.org/10.1080/10573569.2023.2175340

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading

Disabilities? A Meta-Analysis. *Journal of Learning Disabilities*, *51*(1), 73–84. https://doi.org/10.1177/0022219416688170

WOODALL, B. (2010). Simultaneous Listening and Reading in ESL: Helping Second Language Learners Read (and Enjoy Reading) More Efficiently—WOODALL - 2010—TESOL Journal—Wiley Online Library. *TESOL Journal*, *1*(2), 186–205.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

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

www.cambridgeenglish.org. (n.d.). What are the different 'levels' of learning a language? | Cambridge English. [online] Available at: https://www.cambridgeenglish.org/learning-english/parents-and-children/how-to-support-your-child/what-are-the-different-levels-of-learning-a-language/.

Yaacob, H., Zakariya, N. Z., & Mohd Rashid, S. M. (2024). Technology-based Interventions for Dyslexic Children: A Systematic Literature Review (SLR). *International Journal of Academic Research in Business and Social Sciences*, *14*(2), Pages 574-595. https://doi.org/10.6007/IJARBSS/v14-i2/20826

Yali, G. (2010, February 1). *L2 Vocabulary Acquisition Through Reading—Incidental Learning and Intentional Learning. | Chinese Journal of Applied Linguistics (Foreign Language Teaching & Research Press) | EBSCOhost*. https://openurl.ebsco.com/contentitem/gcd:57233898?sid=ebsco:plink:crawler&id=ebsco:gcd:57233898

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, *2*, 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696–725. https://doi.org/10.1177/1362168820913998

Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, *35*(4), 696–742. https://doi.org/10.1080/09588221.2020.1744666

Zhang, Y., & Wu, W. (2021). How effective are lexical richness measures for differentiations of vocabulary proficiency? A comprehensive examination with clustering analysis. *Language Testing in Asia*, *11*(1), 15. https://doi.org/10.1186/s40468-021-00133-6

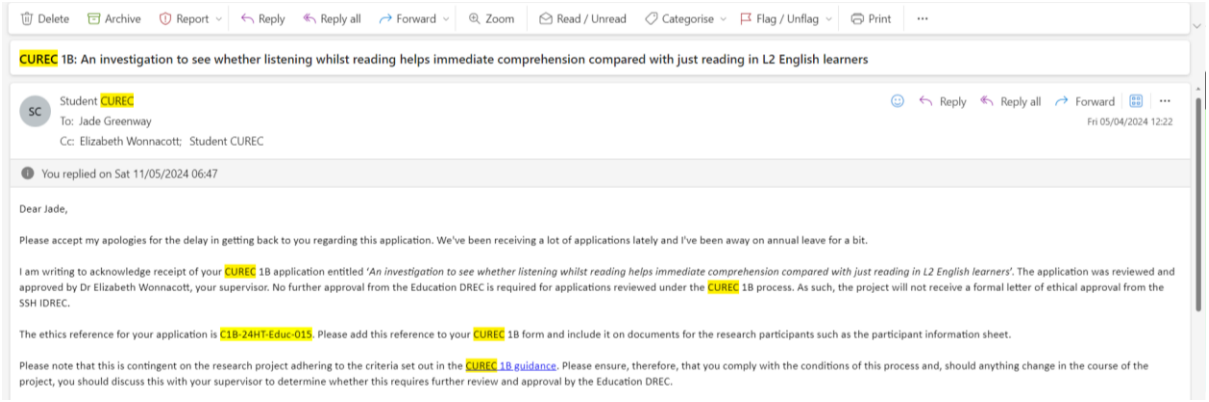
Zhao, A., Guo, Y., Biales, C., & Olszewski, A. (2016). *Exploring learner factors in second language (L2) incidental vocabulary acquisition through reading*. http://hdl.handle.net/10125/66903

Zhao, K., Zhou, J., & Zou, B. (2021). Developing subject knowledge co-construction and specific language use in a technology-enhanced CLIL programme: Effectiveness and productive patterns. *International Journal of Bilingual Education and Bilingualism*, *25*(6), 2172–2185. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1080/13670050.2021.1890688

(2024) أ. س. ع. م. الفل, & ر. ا. ابراهيم. Leveraging Assistive Technology in Language Learning: The Effectiveness of Text-to-Speech (TTS) on improving Phonemic Awareness and Orthographic Knowledge of Dyslexic Children. □□□□□□□□□ □□□□□□ □□□□□□□□ □□□□□□ □□□□ □□□□□ □□□□. https://doi.org/10.21608/musi.2024.290580.1161

Appendices

Appendix 1: CUREC approval



Delete | Archive | Report | Reply | Reply all | Forward | Zoom | Read / Unread | Categorise | Flag / Unflag | Print

CUREC 1B: An investigation to see whether listening whilst reading helps immediate comprehension compared with just reading in L2 English learners

Student CUREC
To: Jade Greenway
Cc: Elizabeth Wonnacott; Student CUREC

Reply | Reply all | Forward
Fri 05/04/2024 12:22

You replied on Sat 11/05/2024 06:47

Dear Jade,

Please accept my apologies for the delay in getting back to you regarding this application. We've been receiving a lot of applications lately and I've been away on annual leave for a bit.

I am writing to acknowledge receipt of your CUREC 1B application entitled '*An investigation to see whether listening whilst reading helps immediate comprehension compared with just reading in L2 English learners*'. The application was reviewed and approved by Dr Elizabeth Wonnacott, your supervisor. No further approval from the Education DREC is required for applications reviewed under the CUREC 1B process. As such, the project will not receive a formal letter of ethical approval from the SSH IDREC.

The ethics reference for your application is C1B-24HT-Educ-015. Please add this reference to your CUREC 1B form and include it on documents for the research participants such as the participant information sheet.

Please note that this is contingent on the research project adhering to the criteria set out in the CUREC 1B guidance. Please ensure, therefore, that you comply with the conditions of this process and, should anything change in the course of the project, you should discuss this with your supervisor to determine whether this requires further review and approval by the Education DREC.

UNIVERSITY OF OXFORD

Central University Research Ethics Committee (CUREC)

CUREC 1B Application form for research projects in the social sciences and humanities with low-risk ethical issues

The University of Oxford places a high value on the knowledge, expertise, and integrity of its members and their ability to conduct research to high standards of scholarship and ethics. The research ethics review process has been established to ensure that research involving human participants is conducted in a way that respects the dignity, rights, and welfare of participants, and minimises risk to participants, researchers, third parties, and to the University itself. It is assumed that all members of the University will take their responsibilities and obligations seriously, and will ensure that their research involving human participants is conducted according to established principles and good practice in their field and in accordance, where appropriate, with legal requirements.

This form is currently being piloted in the following Departments: Asian & Middle Eastern Studies (AMES), Computer Science, Education, Geography and Environment (SoGE), International Development (ODID), Linguistics, Philology, and Phonetics (LPP), Music and Sociology, and is not yet available to researchers in other Departments.

Ethical reference: C1B-24HT-Educ-015

SECTION A: Researchers	
1. Name of researcher or student	Jade Greenway
2. Department or Institute	Education
3. Degree programme, if student research	MSc ALSA
Copy and paste the following four rows as necessary to complete for each additional researcher who will be involved in this study, including student(s).	
4. Name of Principal Investigator (PI) (if different from the answer to A1), student's supervisor or other researchers	Elizabeth Wonnacott
5. Department or Institute	Education
6. Role in research	Supervisor
7. The introductory core research integrity course is compulsory for all University of Oxford research students (either on graduate taught courses or taking research degrees). There is also a refresher course which may be more suitable for experienced University of Oxford researchers. Please confirm that all staff and research students have undertaken either the core or the refresher course, or that undergraduates have received suitable training.	Yes ☒

SECTION B: Filter questions		
This section determines whether this CUREC 1B form is suitable for the research project. Please indicate with an 'X'.	Yes	No

CUREC 1B version 1.0 – pilot MT 2023 – for use in AMES, Computer Science, Education, SoGE, ODID, LPP, Music and Sociology only

Appendix 2: Text 1

Dr Tomas Streyer looked around the control room at his team of scientists and engineers. He was pretending to be calm, but he was both excited and terrified. The next few minutes would be the starting point of years more research towards understanding the secrets of how the universe began.

He looked out of the window at the beautiful blue summer sky and took a deep breath. 'Ready,' he said, and pressed the first button, bringing to life the complicated computers and machines around them. 'Set,' he said, and pressed the second button, switching on the huge particle accelerator that lay in a huge underground laboratory, deep beneath the towns and fields of Switzerland. 'Go,' he said, at exactly twelve o'clock, and pressed the final button. For a second, everything went absolutely black, as if he had gone blind. Tomas cried out in shock, but the lights were already on again. Whatever that was, it was not supposed to have happened. 'Everybody check the systems!' he ordered, but nothing seemed to be wrong with them. The experiment was working, just as he had hoped it would. 'Look outside,' said Tomas's assistant in a frightened voice.

Instead of the perfect summer day of five minutes ago, the sky was darker than the blackest night. The sun disappeared.

People were shouting and screaming, calling their families on the telephone, afraid they had all suddenly disappeared too. Tomas ignored their noise. He sat at the main computer and started reading the data from his experiment. Nothing there explained what was happening. He rushed for the exit, his team running behind him until they were all outside the laboratory building. Everyone else in the research centre was outside, panicking and confused. They were using the screens and torches on their mobile phones to see where they were going, waving them around like giant fireflies. Several people got in their cars and turned on the headlights, driving them to the entrance to make a small area of brightness in the dark for everybody to crowd together. When the street lights turned on, some people cheered, but most of them were still frightened. Then, almost twenty minutes after Tomas had started his experiment, without warning the sun reappeared in the sky, warm and yellow, and the black sky turned blue again. But hours later, when night fell again, no one was celebrating. Because although the moon rose as usual, there was not a star to be seen.

No one wanted to know what Tomas had intended to happen. No one wanted to know what his research had been about. What did that matter? All they cared about was what had happened. He had stolen all the stars – or that's what the newspapers said. And when they put him on trial, that's what they charged him with: stealing the stars.

He pleaded 'Not guilty'.

'So neither you nor your experiment stole the stars from us,' said the lawyer. 'No,' Tomas said. 'You just made it so that we can't see them anymore.' After a long pause, Tomas sighed. 'Yes.' The lawyer raised an eyebrow. 'How is that any different?' Tomas didn't have an answer, not one anyone would understand anyway. He had an idea, but it would take years to prove it. Instead, he changed his plea to 'guilty'.

The trial gave the world someone to blame for what it had lost, but sending Tomas to prison for years wouldn't change anything. Instead, they gave him a fake punishment. Tomas was sentenced to work at the abandoned Extremely Large Telescope in Paranal, Chile.

Each evening, he would watch the sun go down. It vanished below the horizon exactly eight minutes and twenty seconds after it actually descended below the curve of the Earth. The laws of physics

remained the same, much to Tomas's relief. Light travelled at the same speed it always had. He hadn't changed reality that much. His idea was that if light travels at a certain speed maybe the absence of light travels at the same speed. There was no way to prove his belief. So he had to wait.

High in the mountains of Chile, Tomas continued to watch the night. His enormous telescope pointing towards a particular point of the sky, even though it was always just as empty as every other part. For 1,596 black nights – nearly four and a half years – there was no change to the night sky.

But that was OK. It didn't disprove Tomas's idea, it supported it. Tomas thought about the absence of light passing the sun and continuing out towards the edge of space. It would take 1,596 nights to pass the nearest star, and another 1,596 nights for that star's light to reach the Earth again.

And then one night, almost nine years after the disaster, Tomas looked up from his telescope and saw Alpha Centauri twinkling back at him. The first star. He wiped the tears from his eyes and made a wish. Billions of other people's wishes followed right after it.

Appendix 2.1: Text 1 questions

1. Just after the experiment, who could explain what had happened?
    - Thomas
    - Thomas' assistant
    - The lawyer
    - No one
2. What could be seen after 20 minutes?
    - No light at all
    - The sun
    - The moon
    - The stars
3. What was the problem with Thomas' theory?
    - He did not have the right equipment to prove it.
    - It needed time to prove right or wrong.
    - He was not sure it was worth investigating.
    - He did not know how to do the calculations.
4. Why did it take almost nine years to see the first star?
    - Thomas did not look at the sky every night.
    - Thomas was looking at the wrong place in the sky.
    - It took nine years for Alpha Pegasi's light to be seen.
    - It took four and a half years for the darkness to pass Alpha Centauri and four and a half years for its light to reach Earth.
5. Choose the correct word for this gap:
   He sat at the main computer and started ___ the data from the experiment.
    - Read
    - Reading
    - Sight
    - Looks
6. Choose the correct word for this gap:
   Thomas changed his ___ to guilty in the end.

- Plea
- House
- Guess
- Begging

Appendix 3: Text 2

What does your voice say about you?

It's not what you say, it's the way you say it that matters. There is a lot of truth to this statement. Your voice reveals a lot more about you than you think, and other people are quick to make judgements about you on the strength of your voice alone. Scientists have discovered that these judgements are surprisingly similar, which means that when listeners hear a voice, they often agree on the kind of personality traits they expect that person to have. For example, when they hear a person with a nasal voice, the kind of voice that sounds as if someone has a cold, they usually agree they are listening to a bored or complaining person.

It has also been shown that both men and women can judge physical traits such as a man's upper body strength correctly by listening to his voice. It is not clear what characteristics of the voice convey this information and although it is often believed that stronger men have deeper voices, this is not strictly true.

Our voices can also reveal our age. Children have higher-pitched voices than adults and during teenage years the voices of both girls and boys become deeper. In boys, this change is greater and can be very noticeable if it happens suddenly. This dramatic change is caused when the vocal cords which control the voice become longer and thicker. Sometimes a boy may have difficulty controlling his voice as it 'breaks'.

What if your voice gives the wrong impression and stops you communicating as you would like to?

Well, voices can be trained and it is true that some politicians have trained their voices so they sound more powerful. This is perhaps especially important for some female politicians, whose high-pitched squeaky voices could prevent them from creating the serious image that they need in public life. Other professionals, for example, teachers, are in danger of damaging their voices if they shout too much. They could also benefit from learning to use their voices more effectively.

Of course, people rarely want to change their voices completely, even if it were possible. Our voices are a unique part of us. We recognise familiar voices, although sometimes members of the same family have voices that are difficult to tell apart. When you hear a recording of yourself, you may be surprised how similar your voice is to the voices of close relatives. Friends can also speak in similar ways. As they spend time with each other, they often pick up accents without even realising it. This helps friends understand each other better.

Another thing about our voices that may be hard to control is the way they are affected by both our health and our feelings. The voice can reveal how people are feeling both physically and emotionally. A strong voice is usually a sign of health and people will often say that someone 'sounds well' after a telephone conversation, whereas when a person's health is failing, their voice may become quieter and weaker. Emotions can clearly affect our voices, too. It is obviously hard not to raise your voice when you are angry and people who experience strong emotions may actually be speechless! Additionally, your voice can tell someone how insecure or confident you are because of the pitch, but also the rhythm. If you are nervous, our voices tend to speed up and we 'rush through' our words, or we can use irregular pauses which breaks the rhythm and indicates how you are nervous.

Whatever your voice says about you, it is the first impressions that really count. Experiments have shown that as soon as you open your mouth, your listener makes a snap judgement about your personality. When this is positive, the expression, 'You had me at 'Hello'!' is not an exaggeration.

Appendix 3.1: Text 2 questions

1. What is suggested in the first paragraph?
    - Voices are never an accurate indicator of character.
    - Nasal voices always mean someone is suffering from a cold.
    - Different voice styles are perceived to reveal different traits.
    - Words are important.
2. What is true about men's voices?
    - They are always higher pitched than a woman's.
    - We can guess a man's strength based on their voice.
    - Men use less words than women.
    - Other men judge male voices more.
3. How do children's voices change as they grow?
    - All teenagers' voices become deeper.
    - Children have the deep voice.
    - Female voices 'drop' during puberty.
    - Our voices change overnight when we are 10.
4. What did the author mean by "You had me at Hello!"
    - I had to say hello.
    - I would like to meet you.
    - The smile caught my attention first.
    - I liked you when I first heard your voice.
5. Choose the correct word for this gap:
   Of course, people rarely want to change their voices ___, even if it was possible.
    - Slight
    - Completely
    - To make it deeper.
    - To a higher pitch.
6. Choose the correct word for this gap:

In other professions, ___ are in danger of damaging their voices if they shout too much.

- Doctors
- Builders
- Teachers
- Shop managers

Appendix 4: Videos

For the purpose of this appendix, all videos have been uploaded to YouTube and set to unlisted. These links will take you to the page so they can be viewed. However, due to the nature of uploading them on YouTube they cannot be anonymised.

Audio speed questions

-25: https://youtube.com/shorts/vfEanSLVQxA?feature=share

0: https://youtube.com/shorts/IK4w8E8xRG4?feature=share

+25: https://youtube.com/shorts/NgGfy4yIUHE?feature=share

Text 1:

-25: https://youtu.be/2b8rwbcZZsk

0: https://youtu.be/J9LeglxBEYo

+25: https://youtu.be/ET8NtdgMU20

Text 2:

-25: https://youtu.be/PJ0tw72ovZk

0: https://youtu.be/aMhv60rM0NU

+25: https://youtu.be/aMhv60rM0NU

Appendix 5: Demographic questionnaire

There will now be a few questions to answer.

They are all OPTIONAL and just provide some additional data. Thank you.

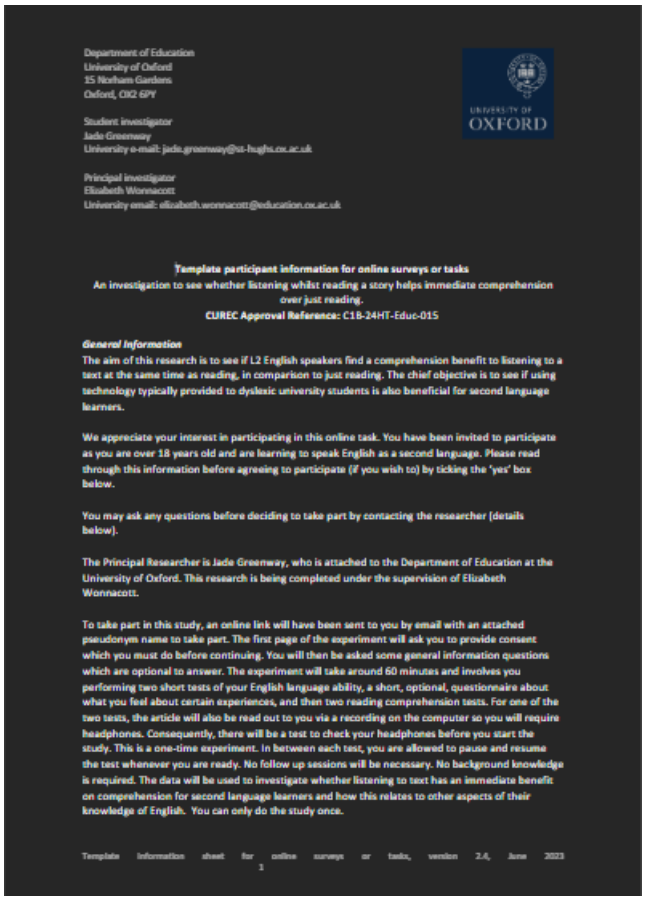
1. What is your Gender?
   - Male
   - Female
   - Prefer not to say
   - Other – Please specify: ____
2. What is your age in years and months?
   Please answer this question with year then month - e.g. 22 years 5 months.
3. Have you ever lived in an English speaking country? If yes, which?
4. How long have you lived there?
   Please write your answer as the number followed by the time: e.g. 8 months or 6 years
5. What languages do you speak?
6. How fluent would you rate yourself in all languages?

If you can, please provide an English level using the image below if needed.

Level			General description	Cambridge English Exam
Proficient user	C2	Mastery	Highly proficient – can use English very fluently, precisely and sensitively in most contexts	Cambridge English: Proficiency
	C1	Effective Operational Proficiency	Able to use English fluently and flexibly in a wide range of contexts	Cambridge English: Advanced
Independent user	B2	Vantage	Can use English effectively, with some fluency, in a range of contexts	Cambridge English: First/First for Schools
	B1	Threshold	Can communicate essential points and ideas in familiar contexts	Cambridge English: Preliminary/ Preliminary for Schools
Basic user	A2	Waystage	Can communicate in English within a limited range of contexts	Cambridge English: Key/Key for Schools Cambridge English: Flyers
	A1	Breakthrough	Can communicate in basic English with help from the listener	Cambridge English: Movers Cambridge English: Starters

Thank you for your help.

Appendix 6: Participant information sheet



Department of Education
University of Oxford
15 Norham Gardens
Oxford, OX2 6PY

Student investigator
Jade Greenway
University e-mail: jade.greenway@st-hughs.ox.ac.uk

Principal investigator
Elizabeth Wonnacott
University email: elizabeth.wonnacott@education.ox.ac.uk

Template participant information for online surveys or tasks

An investigation to see whether listening whilst reading a story helps immediate comprehension over just reading.

CUREC Approval Reference: C1B-24HT-Educ-015

General Information

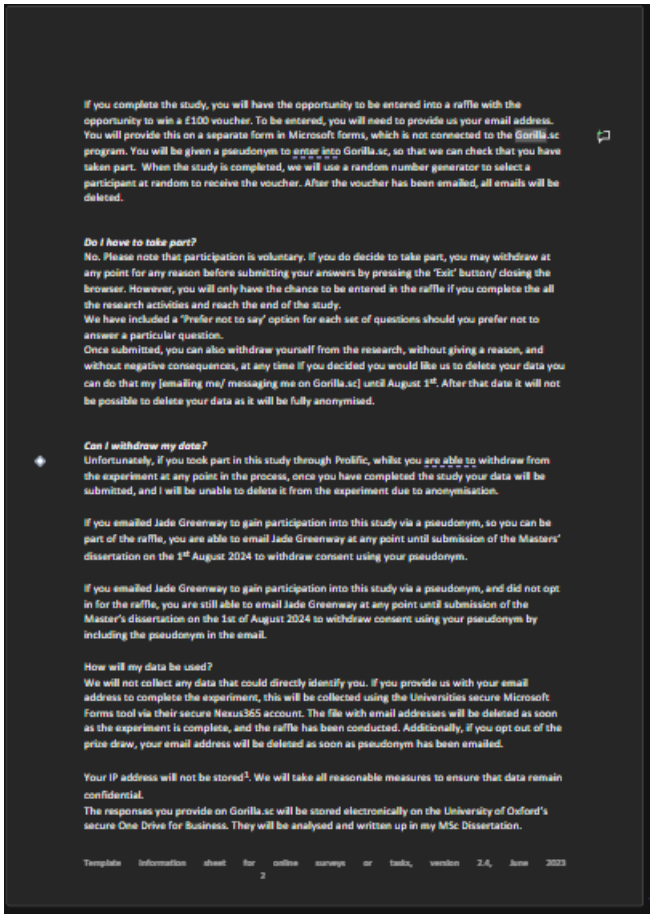
The aim of this research is to see if L2 English speakers find a comprehension benefit to listening to a text at the same time as reading, in comparison to just reading. The chief objective is to see if using technology typically provided to dyslexic university students is also beneficial for second language learners.

We appreciate your interest in participating in this online task. You have been invited to participate as you are over 18 years old and are learning to speak English as a second language. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below.

You may ask any questions before deciding to take part by contacting the researcher (details below).

The Principal Researcher is Jade Greenway, who is attached to the Department of Education at the University of Oxford. This research is being completed under the supervision of Elizabeth Wonnacott.

To take part in this study, an online link will have been sent to you by email with an attached pseudonym name to take part. The first page of the experiment will ask you to provide consent which you must do before continuing. You will then be asked some general information questions which are optional to answer. The experiment will take around 60 minutes and involves you performing two short tests of your English language ability, a short, optional, questionnaire about what you feel about certain experiences, and then two reading comprehension tests. For one of the two tests, the article will also be read out to you via a recording on the computer so you will require headphones. Consequently, there will be a test to check your headphones before you start the study. This is a one-time experiment. In between each test, you are allowed to pause and resume the test whenever you are ready. No follow up sessions will be necessary. No background knowledge is required. The data will be used to investigate whether listening to text has an immediate benefit on comprehension for second language learners and how this relates to other aspects of their knowledge of English. You can only do the study once.



If you complete the study, you will have the opportunity to be entered into a raffle with the opportunity to win a £100 voucher. To be entered, you will need to provide us your email address. You will provide this on a separate form in Microsoft forms, which is not connected to the Gorilla.sc program. You will be given a pseudonym to enter into Gorilla.sc, so that we can check that you have taken part. When the study is completed, we will use a random number generator to select a participant at random to receive the voucher. After the voucher has been emailed, all emails will be deleted.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser. However, you will only have the chance to be entered in the raffle if you complete the all the research activities and reach the end of the study.

We have included a 'Prefer not to say' option for each set of questions should you prefer not to answer a particular question.

Once submitted, you can also withdraw yourself from the research, without giving a reason, and without negative consequences, at any time if you decided you would like us to delete your data you can do that my [emailing me/ messaging me on Gorilla.sc] until August 1st. After that date it will not be possible to delete your data as it will be fully anonymised.

Can I withdraw my data?

Unfortunately, if you took part in this study through Prolific, whilst you are able to withdraw from the experiment at any point in the process, once you have completed the study your data will be submitted, and I will be unable to delete it from the experiment due to anonymisation.

If you emailed Jade Greenway to gain participation into this study via a pseudonym, so you can be part of the raffle, you are able to email Jade Greenway at any point until submission of the Masters' dissertation on the 1st August 2024 to withdraw consent using your pseudonym.

If you emailed Jade Greenway to gain participation into this study via a pseudonym, and did not opt in for the raffle, you are still able to email Jade Greenway at any point until submission of the Master's dissertation on the 1st of August 2024 to withdraw consent using your pseudonym by including the pseudonym in the email.

How will my data be used?

We will not collect any data that could directly identify you. If you provide us with your email address to complete the experiment, this will be collected using the Universities secure Microsoft Forms tool via their secure Nexus365 account. The file with email addresses will be deleted as soon as the experiment is complete, and the raffle has been conducted. Additionally, if you opt out of the prize draw, your email address will be deleted as soon as pseudonym has been emailed.

Your IP address will not be stored[1]. We will take all reasonable measures to ensure that data remain confidential.

The responses you provide on Gorilla.sc will be stored electronically on the University of Oxford's secure One Drive for Business. They will be analysed and written up in my MSc Dissertation.

Template information sheet for online surveys or tasks, version 2.4, June 2023
2

Research data will be stored for at least three years after publication or public release of the work of the research. I may also share fully anonymized data on public repositories such as the Open Science Framework.

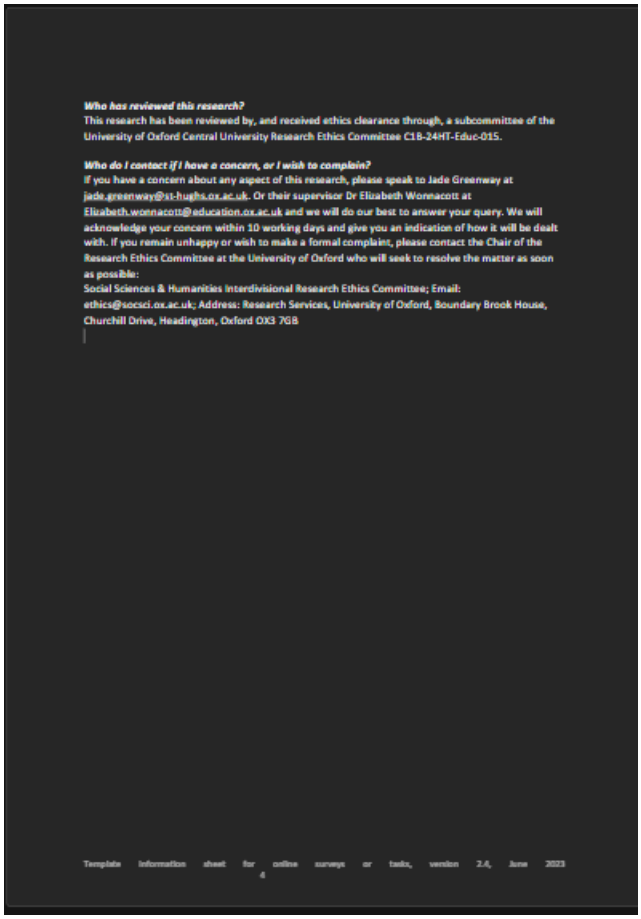
What will happen to me if I take part in the research?
If you agree to take part in this experiment, you will take part in a study performed on Gorilla.sc.sc. The first page of the experiment will ask you to provide consent which you must do before continuing. The experiment will take around 60 minutes and involves you performing two short tests of your English language ability, and then two reading comprehension tests. For one of the two tests, the article will also be read out to you via a recording on the computer so you will require headphones and there will be a test to check headphones before you start the study. This is a one-time experiment. In between each test, you are allowed to pause and resume the test whenever you are ready. No follow up sessions will be necessary.

Who will have access to my data?
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 we perform in the public interest. Further information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.
The results will be written up for an MSc degree.

What information will be collected and why is the collection of this information relevant for achieving the research objectives?
Other than your performance on the two English tests and the two English reading and reading-while-listening comprehension tasks, we will collect answers to the following questions:

1. What languages you speak – including English.
2. How old you are.
3. How long you have been learning English, and to self-rate your proficiency.
4. Are you currently living in an English-speaking country or have you ever lived in and English-speaking country, and for how long
5. Your gender (if you chose to give this)

These questions are relevant to achieving the research objectives because it allows us to understand factors about our participants which are relevant for understanding their performance in our tasks.
Within the data you will be identified using a pseudonym that we will give you. [In addition, during the experiment your name, pseudonym and email address will be entered into a linkage file that will be stored on the University of Oxfords secure Nexus OneDrive account, password protected, separately from the data if you would like to be included in the prize draw. At the end of the experiment, a participant will be randomly selected to receive the prize, using a random number generator. They will be emailed the amazon voucher. At this point, the linkage file will be deleted. Or, if you decide not to enter the prize draw, your name and all identifiable information will be deleted as soon as my reply is sent. No identifiable names or email addresses will be used in the data. All emails will be deleted once the experiment is complete.]
The participants name and email are only gathered to allow for one entry into the prize draw, limiting participants retaking the experiment.



Who has reviewed this research?
This research has been reviewed by, and received ethics clearance through, a subcommittee of the University of Oxford Central University Research Ethics Committee C1B-24HT-Educ-015.

Who do I contact if I have a concern, or I wish to complain?
If you have a concern about any aspect of this research, please speak to Jade Greenway at jade.greenway@st-hughs.ox.ac.uk. Or their supervisor Dr Elizabeth Wonnacott at Elizabeth.wonnacott@education.ox.ac.uk and we will do our best to answer your query. We will acknowledge your concern within 10 working days 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:
Social Sciences & Humanities Interdivisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

Template information sheet for online surveys or tasks, version 2.4, June 2023
4

Please note this was edited depending on whether the information sheet was for the participants gained via the schools or Prolific.

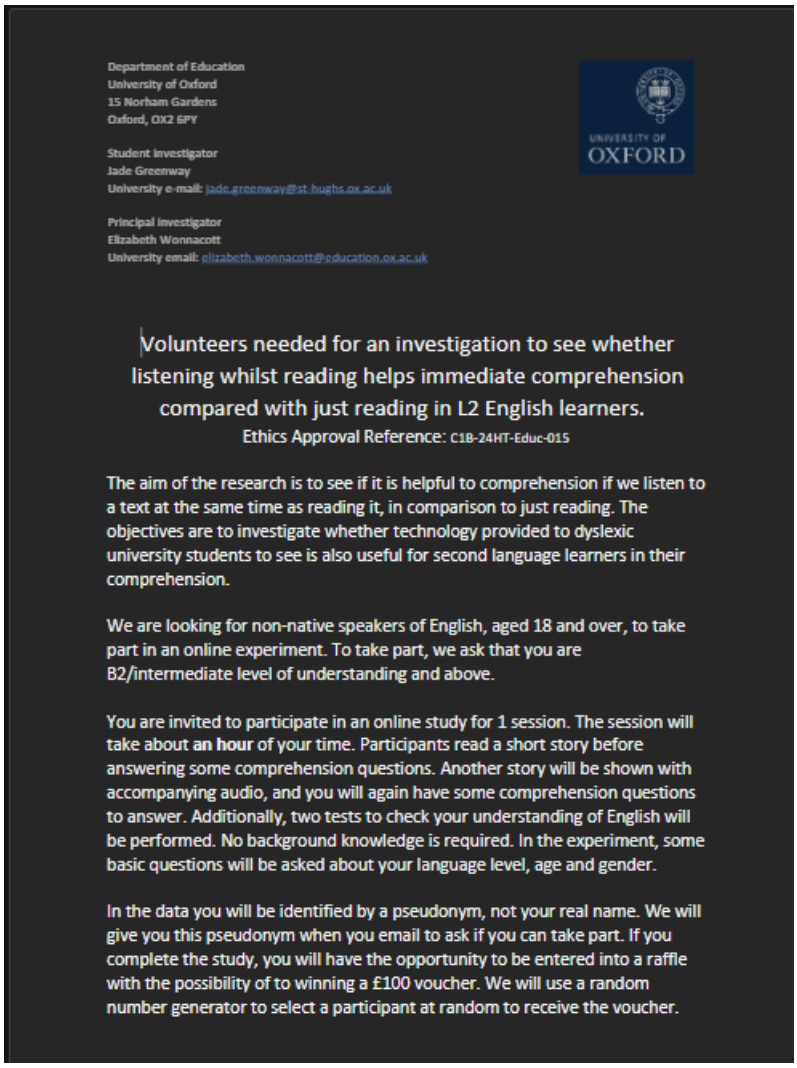
Appendix 7: Consent form on Gorllia.sc

Fill in your consent question here

	I Agree	I Disagree
Please note that you may only participate in this survey if you are 18 years of age or over.	☐	☐
If you have read the information above and agree to participate with the understanding that the data you submit will be processed accordingly, please tick the box below to start.	☐	☐

Next

Appendix 8: Recruitment poster



Department of Education
University of Oxford
15 Norham Gardens
Oxford, OX2 6PY

Student investigator
Jade Greenway
University e-mail: jade.greenway@st-hughs.ox.ac.uk

Principal investigator
Elizabeth Wonnacott
University email: elizabeth.wonnacott@education.ox.ac.uk

UNIVERSITY OF OXFORD

Volunteers needed for an investigation to see whether listening whilst reading helps immediate comprehension compared with just reading in L2 English learners.

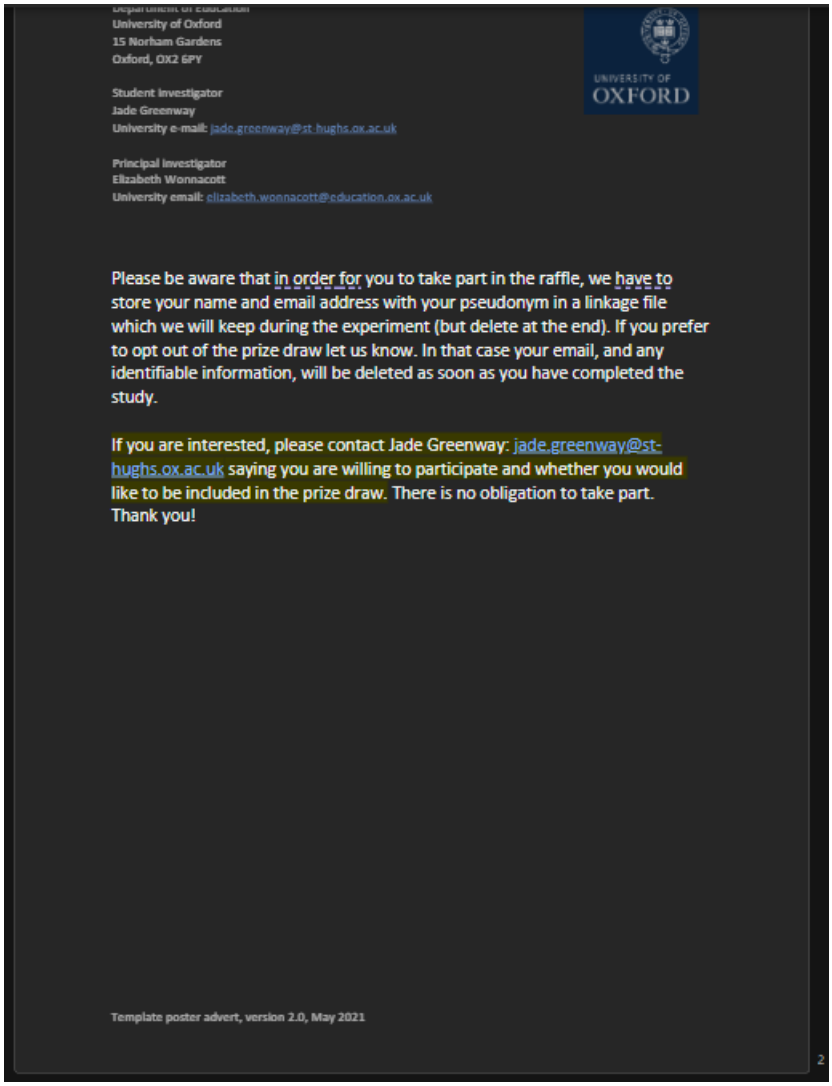
Ethics Approval Reference: C1B-24HT-Educ-015

The aim of the research is to see if it is helpful to comprehension if we listen to a text at the same time as reading it, in comparison to just reading. The objectives are to investigate whether technology provided to dyslexic university students to see is also useful for second language learners in their comprehension.

We are looking for non-native speakers of English, aged 18 and over, to take part in an online experiment. To take part, we ask that you are B2/intermediate level of understanding and above.

You are invited to participate in an online study for 1 session. The session will take about **an hour** of your time. Participants read a short story before answering some comprehension questions. Another story will be shown with accompanying audio, and you will again have some comprehension questions to answer. Additionally, two tests to check your understanding of English will be performed. No background knowledge is required. In the experiment, some basic questions will be asked about your language level, age and gender.

In the data you will be identified by a pseudonym, not your real name. We will give you this pseudonym when you email to ask if you can take part. If you complete the study, you will have the opportunity to be entered into a raffle with the possibility of to winning a £100 voucher. We will use a random number generator to select a participant at random to receive the voucher.



Appendix 9: Dyslexia questionnaire

This came after the sentence verification task and participants were provided scaled slider with rarely on one side and always on the other.

I now have a short questionnaire for you to complete. Feel free to leave questions blank if you would like. Thank you.

1. Do you confuse visually similar words such as cat and cot?
2. Do you lose your place or miss out lines when reading?
3. Do you re-read paragraphs to understand them?
4. Do you get confused when given several instructions at once?
5. Do you find it difficult to find the right word to say?
6. How easy do you find it to sound out words such as e-le-phant?
7. When writing, do you find it difficult to organise thoughts on paper?
8. How easy do you find it to recite the alphabet?
9. How hard do you find it to read aloud?
10. Did you learn your multiplication tables easily?

Appendix 10: Email template to schools

To whomever this may concern,

I hope you are well. My name is Jade Greenway and I am a Master of Science student at the Department of Education at University of Oxford.

I am working on my dissertation at the moment, and I am wondering if you would be willing to forward my experiment in the coming month to some of the English Learners at your school.

The experiment is performed online using the platform Gorilla.sc. It is an investigation to see if using text-to-speech software has an immediate benefit for English language learners. There will be 2 tests to see what their language understanding is like, and then 2 comprehension tests based on articles. In total, the experiment should take about an hour for the participants. There will be no identifiable data collected from the participants. Whilst they are able to withdraw from the experiment at any point in the process, once the students have submitted the data at the end of the intervention via Gorilla.sc, I will be unable to withdraw the data from the experiment due to the lack of identifiable information. I will make this clear to the participants on the test before they take part. My hope is for the participants to be aged 18 and over with a language level of around B2.

If you have a concern about any aspect of this research, please contact me at jade.greenway@st-hughs.ox.ac.uk. Or my supervisor Dr Elizabeth Wonnacott at elizabeth.wonnacott@education.ox.ac.uk and we will do our best to answer your query. I will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

Thank you for your time, I hope you would be willing to help me out with this dissertation.

Kind regards,

Jade Greenway