

Xuan Wang* and Elsa Tragant

The effect of written text on comprehension of spoken English as a foreign language: A replication study

<https://doi.org/10.1515/iral-2018-0350>

Abstract: The use of written text has been acclaimed to enhance L2 listening comprehension, yet some argue that using written text does not effectively prepare learners to listen in real situations. Thus, the study was conducted to explore the effect of written text on learners' perceived difficulty, listening comprehension and learning to listen through replicating the research by Diao, Chandler & Sweller (2007. The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology* 237–261). Participants were 101 low-proficient English learners who were divided into three groups: listening with subtitles, listening with a full script and listening only. Each group first listened to a passage in their respective mode, then all three groups listened to another passage in the listening-only mode. Participants rated their perceived difficulty and completed a free recall task after each listening. Results suggest that the difficulty of written text should be tuned with learners' proficiency level so that they can benefit from the presence of written text in listening.

Keywords: second language listening, cognitive load theory, dual-coding theory, second language teaching, replication studies

1 Introduction

Listening, as a fundamental language skill, refers to how individuals receive, attend to and assign meaning to the aural stimuli (Wolvin and Coakley 1985). In the realm of second language acquisition (SLA), listening has a central status as it contributes to learners' understanding of the target language and development of the other three language skills (Vandergrift 2007). In recent years, greater

*Corresponding author: Xuan Wang, Department of English and German Philology, Universitat de Barcelona, Barcelona, Catalunya, Spain, E-mail: wxinuk@icloud.com

Elsa Tragant, Department of English and German Philology, University of Barcelona, Barcelona, Spain, E-mail: tragant@ub.edu

emphasis has been placed on listening skills in many language programmes (Field 2008; Vandergrift and Goh 2012) because there are quite a few challenges for L2 learners to improve and for instructors to help with (Siegel 2013). To illustrate, it is possible that some instructors are little familiar with the intricacies of listening skills because most people acquire L1 listening skills without obvious efforts (for example see a list of misconceptions about the teaching of listening in Brown and Lee 2015: 318). Consequently, less thought is given to how listening skills develop in L2 learners, and traditional product-focused teaching methods, which do not suffice to cultivate learners' listening abilities, remain in many L2 classes (Graham 2017).

Accordingly, researchers have been conducting studies to explore effective and efficient ways of teaching listening skills (Vandergrift 2007). In the English as a Foreign Language (EFL) context, listening materials are the main sources of input for learners. According to Krashen (1985), such input should be comprehensible to optimise language acquisition. One common procedure is the use of the written text, such as subtitles¹ and scripts,² alongside audio to make it comprehensible and enhance learners' intake. There are some studies acclaiming the positive effects of written text on listening comprehension (e. g. Chang 2009; Chang and Millett 2014), yet results of existing research in this area are still inconclusive and even contradictory (Vanderplank 2016). At the heart of this argument are two main theories: the cognitive load theory (Sweller 1988) and the dual-coding theory (Paivio 1990), both of which will be further discussed in the present study. The cognitive load theory was the theoretical basis of Diao et al. (2007) which the present study intends to replicate with an aim to explore the effects of the written text in L2 listening on learners' perceived difficulty, comprehension of spoken English and their learning to listen.

2 Literature review

2.1 Cognitive load theory

Sweller's (1988) cognitive load theory (CLT) provides a theoretical framework linking human cognition and instructional design. The theory examines how

¹ Subtitles in this thesis refer to the verbatim on-screen L2 texts accompanying the L2 sound as the study by Diao et al. (2007).

² The script is a single large segment of the whole L2 written text identical to the listening material; in this thesis, listening with a full script is regarded as listening while reading.

cognitive resources are focused and used when processing and learning information from an evolutionary view of the human cognition architecture (Sweller 1988, Sweller 2010; Paas et al. 2003; Sweller et al. 2011; Paas and Sweller 2012). CLT assumes the human cognitive system has a limited working memory capacity and is concerned with the relationships between working memory and long-term memory as well as their effects on learning and problem solving (Sweller 2011). According to Geary (2008), there are two types of knowledge: biologically primary knowledge and secondary knowledge. Humans are specifically evolved to acquire biologically primary knowledge in an effortless, automatic, and unconscious way without any instruction, such as learning to listen and speak the mother tongue (Tricot and Sweller 2014). By contrast, biologically secondary knowledge, mainly domain-specific knowledge, requires the assistance of explicit instruction and conscious efforts (Sweller et al. 2007). CLT exclusively focuses on the secondary knowledge, such as the learning of a second language in educational or training contexts because processing and learning secondary information are directly related to working memory limitations (Paas and Sweller 2012). In the last decade, CLT has increasingly relied on the biologically evolutionary view, which treats the human cognition system as an analogy to other natural information processing systems, by natural selection. According to CLT, human cognition is comprised of five main principles³ (Sweller 1994, Sweller 1999, Sweller 2004, Sweller 2009, and Sweller 2016).

As proposed by these five principles, long-term memory has large capacity and duration, and comprises a large amount of domain-specific, relatively permanent information (Groot et al. 1996). Long-term memory is the driving force of most human cognitive action and activities, which functions as an information store; it is central to different areas of human endeavour because information stored in long-term memory can contribute to the derivation of problem-solving skills. The contents of a person's long-term memory are mostly borrowed from the long-term memory of others, through imitating what they do, listening to what they say and reading what they write (Bandura 1986). Such information will be reorganized and restructured to be stored in long-term memory as schemas, which are cognitive constructs that categorize and incor-

³ The five principles of CLT are 1st *Long-term memory and the information store principle*, 2nd *Schema theory and the borrowing principle*, 3rd *Problem solving and the randomness as genesis principle*, 4th *Working memory and the narrow limits of change principle*, 5th *Relationships between long-term and working memory and the environmental organizing and linking principle*.

porate multiple elements of information into a single framework; each schema has a specific function to deal with specific information. Yet there are cases, when knowledge is unavailable for borrowing from others, humans are required to create new knowledge through problem-solving and the randomness as genesis principle. Problem-solving contains the random generation of moves with tests to ascertain effectiveness. In other words, when facing novel problems, individuals cannot simply retrieve the required moves from the long-term memory store. Instead, a possible move is randomly chosen and tested so as to jettison ineffective moves and retain the effective ones in long-term memory for future use. Such novel information is processed by the random generation and test of moves in working memory, which acts as a conduit linking the information in long-term memory and the external environment. However, the working memory is severely constrained in capacity and duration when dealing with new secondary knowledge. Thus, if working memory is burdened by too much load, it will not be effective at information processing. The narrow limits of working memory can be eliminated as people can take signals and cues from their environment and transfer information from long-term memory to working memory to generate and direct environmentally-appropriate actions. During this procedure, the working memory load can be reduced through schema construction and automation.

The cognitive load is the working memory resource that is needed for a certain learning task, and it interferes with the learning process (Sweller and Chandler 1994). Considerable research on CLT is centred around the instructional conditions which can keep the cognitive load of the learning tasks within beneficial limits based on a wide range of cognitive load effects, such as the redundancy effect, the attention-split effect (for a review, see Sweller et al. 2011). The present study focuses on the redundancy effect, which appears when learners have to deal with types of information that are identical but have different surface structures and can be understood in isolation (Sweller and Chandler 1994; Kalyuga et al. 2004; Sweller et al. 2011). The occurrence of unnecessary and redundant information will be detrimental to learning (Chandler and Sweller 1991, Chandler and Sweller 1996) because certain working memory resources must be given to distinguishing which element is necessary, and this is particularly relevant in the case of low-level learners (Leslie et al. 2012). There are different types of redundancy effect, such as the simultaneous presentation of verbal and nonverbal elements (e. g. diagrams with text), and of visual and auditory verbal elements (e.g. written & spoken materials). In accordance with CLT, irrelevant and redundant information will impose an extraneous cognitive load by arousing competition between different cognitive

resources, thus such information should be omitted for the sake of effective learning.

For many years, CLT has been used to interpret the results of many studies in the EFL listening context. For instance, Kalyuga et al. (2004) conducted a study with two groups of participants who were asked to comprehend listening material in two modes: one with the concurrent presentation of identical spoken and written text and the other with spoken text only. Results showed that the comprehension exhibited by the second group was higher than that of the first; this could be attributed to the disadvantages of the redundancy effect. Following this line of inquiry, Diao and Sweller (2007) examined whether the written presentations concurrent with spoken presentations were good for the learners' comprehension, and they concluded that the simultaneous presentation of these two modes was less effective. A more recent study by Mayer et al. (2014) explored the effects of videos and on-screen subtitles on non-native English speakers' comprehension. While it found a positive effect to using videos, adding subtitles did not improve participants' performance on the comprehension tests. In addition, according to two recent reviews (Mayer and Fiorella 2014; Kalyuga and Sweller 2014) on the redundancy effect, materials that contain redundant information are likely to negatively influence learning.

However, even though the above-stated empirical results on CLT were not positive towards the use of written text in support of oral input, the written text is still widely used in L2 classroom procedure. In fact, there is more scholarly literature acclaiming its positive effects on L2 listening rather than negative effects. For example, Vandergrift (2004) proposed an integrated model with a stage for listening with written support, and, more recently, Siyanova-Chanturia and Webb (2016) and Chang (2016) also advocated reading to assist listening. At the heart of these studies is the theoretical framework: dual coding theory by Paivio (1990).

2.2 Dual coding theory

Dual coding theory (DCT) was first proposed by Paivio (1990) to explain the dynamic associative processes of verbal and nonverbal representations in human cognition. Developed from the Morton's Logogen Model (1969), DCT claims that there is a verbal system, which deals with linguistic stimuli, and a non-verbal system, which deals with non-linguistic stimuli. Its underlying assumption is concerned with the effects of redundant information and the associations between the two systems. That is, although the verbal and

non-verbal systems function for different stimuli, they will interact with each other, thereby constructing a cooperative system to recognize, process, restore and recall this information (Clark and Paivio 1991). In today's multimedia educational environment, DCT has been applied as a theoretical framework in the investigation of using video, audio and written text in listening classes.

Several theories and hypotheses are derived from DCT to inform the teaching and research of listening skills, and bi-modal input is one of these; bi-modality posits that the concurrent presentations of aural and orthographic stimuli will lead to better learning outcomes (e.g. Bird and Williams 2002; Charles and Trenkic 2015). It specifically suggests that the simultaneous presentation of two modes of input (e.g. spoken & written) will improve learners' comprehension of the material. Indeed, abundant research has approved the use of written text in enhancing L2 learners' comprehension of spoken language; it has also been seen to increase learners' motivation and attention, and reduce anxiety (Vanderplank 1988). For instance, Bird and Williams (2002) investigated the effects of subtitles by looking at how bi-modal input presentation affected implicit and explicit memory in the learning process; Their findings confirmed the interaction of aural and visual processing systems and suggested that the use of subtitles contributed to more in-depth processing and led to better comprehension by both native and non-native university students. In the study by Moreno and Mayer (2002), the authors compared the presentation of written text supported by audio to the sole presentation of auditory mode. They found higher levels of comprehension of the material among the learners in the former mode. Winke et al. (2010) employed a mixed methods approach to analysing the effects of subtitles, they also identified the facilitative role played by subtitles in L2 learners' listening performance. Similarly, Charles and Trenkic (2015) focused on the abilities of L2 learners to segment speech in listening. Their data revealed the learners in the bi-modal (i.e. spoken text & written text) group enhanced their speech segmentation abilities more than the single modal (i.e. spoken text) group, which was also in line with the positive effects of subtitles on listening abilities. According to the meta-analysis study by Perez et al. (2013), based on 18 studies of subtitling, a significantly large and facilitative effect of subtitles on listening comprehension was confirmed.

In addition to the use of subtitles in listening, some recent studies have investigated the simultaneous presentation of a full script accompanying the auditory material, commonly referred to as listening while reading. However, the results seem to be inconclusive. For instance, Chang and Millett (2014)

carried out a study on the L2 listening fluency development through listening to 10 audio graded readers (10 hours of audio over a period of 13 weeks) among 113 lower-intermediate EFL learners. This study proved that more significant outcomes were produced in the mode of listening while reading rather than the listening-only or reading-only modes. Similarly, Kartal and Simsek (2017) investigated the use of audiobooks among university learners of English studies by looking at how reading while listening to two novels (7 hours of audio over a period of 13 weeks) influence their listening comprehension and attitudes. The results also suggested that students could benefit from using audiobooks in terms of listening comprehension, pronunciation and motivation. Apart from these encouraging findings, less positive findings come from two studies (Chang 2009; Tragant et al. 2016) that include participants with a lower level of proficiency. In the case of Chang (2009), the reading-while-listening mode was compared with the listening-only mode to see their respective effects on the comprehension of two short stories (20 minutes of audio on one seating). In this study, the differential effects of the two input modes were only medium which, as explained by the author, indicated some limitations of reading while listening to guarantee a deeper level of comprehension. In the same study, Chang compared any difference between low- and high-level learners and she found the higher-level learners were equipped with more advanced reading skills and thus they could benefit more from reading the script in global comprehension than the lower-level learners. Similar findings come from Tragant et al.'s study with 56 primary school EFL learners who were enrolled in a reading-while-listening intervention programme over a period of one year. Results showed that these low-level learners' listening comprehension skills improved as much as those from a parallel group involved in a reading only intervention programme.

Based on the reviewed literature, we can see that there is some inconclusiveness in the existing scholarship. Listening with written text is not generally supported by studies that follow the CLT framework and it is only partially supported by those that follow the dual-theory framework. Nevertheless, the procedure is not infrequent among some L2 teachers. To address this apparent contradiction, an experimental computer-based study, which we are going to replicate in this work, was conducted by Diao et al. in (2007); it explored the effects of written text on learners' understanding of spoken English in the EFL context. The participants, who were university students from China with 7 years of English study,⁴ were divided into three groups and were directed to perform

4 There are no other indications of their English proficiency in the original article.

two listening tests. In the first listening each group was exposed to a different mode of presentation: listening-only, listening with simultaneous subtitles and listening with a full script. For the listening-only group, the auditory component was the only component of instruction. For the listening with subtitles group, each subtitle consisted of a single sentence that appeared on the screen while the relevant sentence was spoken and for the listening with a full script group, the student had access to the full script while the participants listening. Results showed that written text that accompanied the listening in the subtitles and script groups facilitated comprehension to a greater extent than the listening only mode. However, in the second listening, when a different procedure was followed that consisted in the three groups being exposed to the same condition (listening only), the listening-only group outperformed the other two groups. This experiment was conducted again in the study with simpler texts and the same results were obtained, which can be interpreted as evidence that the listening only mode facilitated learning to listen to a greater degree than the other two modes. In other words, listening with written text (in this case the presence of a script or subtitles) did not appear to ‘help the construction or automation of schemas relevant to listening comprehension’ (p. 251). In addition, as regards learners’ perceived difficulty of the listening material, in the first listening, the groups with written text reported to produce lower mental load than the listening only group, while in the second listening when all groups listened only with auditory material, no significant difference were found. Given this counterintuitive finding, we set out to replicate the study with a similar population to that of the original study.

Following the approximate replication paradigm, the study has the same purpose as the original study with a very similar design and will thus contribute to the self-correcting nature of scientific inquiry (Abbuhl 2011; Porte 2015). Therefore, in accordance with Diao et al. (2007), the main purpose of the present study was to explore the effects of written text on EFL learners’ perceived listening difficulty, their comprehension of spoken English and the development of their listening abilities. The study was guided by the following three research questions.

- 1) Does the use of written text have an effect on L2 learners’ perception of difficulty in the English Listening?
- 2) To what extent does the use of written text influence L2 learners’ comprehension of spoken English?
- 3) To what extent does the use of written text influence L2 learners’ learning to listen?

3 Methodology

3.1 Research design

3.1.1 Three groups of EFL learners

Being a replication study of Diao et al. (2007), the present investigation included three groups. Participants were divided into three groups as the original study, and on the first listening test, each group was exposed to their respective presentation mode: listening with simultaneous subtitles (Listening + subtitles group), listening with a full script (Listening + script group) and the listening only (Listening + only group). As in Diao et al. (2007), students in this study read the same text, but the former group was presented with the sentences when they listened to them, those in the latter had the full text at their disposal while they listened to the audio. The Listening + only group, as its name suggests, listened without any written components. The auditory component was the same for the three groups.

3.1.2 Three instructional phases

There are one preliminary phase and three main instructional phases in the present study (Table 1). In the preliminary phase, a background information questionnaire was administered (See the Appendix One). There was no such a

Table 1: Research design & data collection procedure.

	Procedure	Timing
Preliminary	1. Background information questionnaire	Approximately 10 mins
Phase-1	2. Teaching of the 10 target words	Approximately 20 mins
	3. Translation Test of the 10 target words	
Phase-2	4. Listening 1-Coffee trees (1 st time)	Approximately 30 mins
	5. Subjective mental ratings of difficulty	
	6. Listening 1-Coffee trees (2nd time)	
	7. Free recall writing task	
Phase-3	8. Listening 2-Roses (1 st time)	Approximately 30 mins
	9. Subjective mental ratings of difficulty	
	10. Listening 2-Roses (2nd time)	
	11. Free recall writing task	

preliminary phase in Diao et al. (2007), and we added this phase in order to obtain participants' personal data (e. g. age, gender, self-report English level). In the first instructional phase (Phase 1), participants in each group were required to learn ten keywords (See the Appendix Two) selected from the listening materials they would be exposed to in phases 2 and 3. The rationale behind this phase was, as Diao et al. (2007) indicated, that lexical ignorance was the main obstacle to listening comprehension. A translation task was then used to check if they had learned these ten words by asking students to translate them into Chinese. As we considered the purpose of this phase to let learners learn these key words, we did not teach them as the original one in different modes (i. e. Listening only versus Listening with written text), instead, all three groups learned these key words in the same order: pronunciation, explanation and two sample sentences and these were all shown on the computer screens. In the second instructional phase (Phase 2), participants needed to listen to an expository passage (Coffee trees) in their corresponding mode twice and were asked to rate how difficult they perceived to understand the material after the first listening in a 9-point Likert Scale, with 1 being extremely easy and 9 being extremely difficult (Rating of difficulty). Then the listening was played for a second time, after which participants were asked to write down as much as they could recall about the passage in English (Recall task). We adopted the free recall task as a measure of listening comprehension and did not use the multiple-choice exercise due to the practical consideration of the time limit in a standard class. The purpose of the third phase (Phase 3) was to check whether the written text would improve or hinder the learners' learning to listen. In this phase, participants were asked to listen to another similar expository passage (Roses). This time the passage was presented solely in an auditory mode to all groups without any written components. Same as Phase 2, a rating of perception in difficulty followed the first listening and a free recall task followed the second listening.

3.2 The listening materials

As mentioned above, there are two expository passages (See the Appendix Four) being used, namely Coffee Trees (196 words) and Roses (215 words). These two were selected from the second experiment in the original study because the other two in the first experiment were found to be difficult for the participants in Diao et al. (2007), thus we used these two passages with a thought that our participants were from the vocational college and the materials should not be difficult for them; we also got the confirmation from the teachers of the potential

participants that the materials from the second experiment were suitable to be used with the students in the present study. These two passages share similar syntax containing the same ten keywords; we further tested the readability index which showed that they were at the same difficulty level, and that learners at or above Grade 6 will find it easy to read and understand.

3.3 Participants

Participants were 101 freshmen majoring in Computer Science from a vocational college in southwestern China. As summarized in Table 2, there were 34 students in the Listening + only group, 45 in the Listening + subtitles group and 22 in the Listening + script group; and these three groups were from three intact classes. The mean age of the participants was 19.06 and they had a similar age onset of English ($M = 10.55$). At the commencement of the study, they had been learning English for more than 8 years. However, according to their self-report data in a four-point Likert scale (with 1 being poor and 4 advanced), their level of English was low since their average response on a four-point Likert scale is 1.68.

Table 2: Characteristics of participants.

	Listening + only (n = 34)	Listening + subtitles (n = 45)	Listening + script (n = 22)
Mean Age	19.30	18.93	18.95
Gender	M = 30; F = 4	M = 37; F = 8	M = 17; F = 5
AOE*	10.61	10.56	10.50
YEL**	8.41	8.38	8.45
SREL***	1.71	1.65	1.66

*Age onset of English; **Years of English Learning; ***Self-report English Level

3.4 Data collection and analysis

As presented in Table 1, the implementation of the three phases of the study together took approximately 90 minutes. Due to technical issues, we did not use the computer programme as in the original study, instead, PowerPoint files inserted with audios of the key words were used for the first phase, and in Phase 2, one video with a full script and another with subtitles were played respectively in the Listening + script and Listening + subtitles groups, while for the Listening + only group, only the audio was played. In the last phase, the

same audio file without written text was used for all groups. Printed answer sheets were prepared for the students, and their teacher was in charge of going through all the phases of the study.

As for the data calculation and data entry, one point was given if participants translated the word with the correct and identical meaning in the passage; and the free recall tasks were scored following the scoring criteria (See Appendix 5) in Diao et al. (2007), so regardless of the grammatical mistakes, writing down one main unit was given one point. In the present study, the main units were calculated and discussed by two researchers and inter-rater reliability was computed with SPSS. Two different researchers first scored the number of main units of the three groups independently, and then *Pearson* correlation test was performed between the two scorings: the results showed a high level of inter-rater reliability and consistency of the scoring, with significant values up to 0.926 for scoring of Coffee trees and 0.940 for Roses.

Preliminary analyses were conducted to explore whether the scores in the translation task and free recall task were normally distributed. Based on the results from the Shapiro-Wilk test of normality and the left-skewed histograms, we decided to use non-parametric tests to compare the group differences and set the significance level at 0.01 throughout the study as a result of multiple tests on the same sample. Non-parametric tests were adequate given the fact that the three groups differed significantly in terms of their size. Therefore, a Kruskal-Wallis Test was used to compare the group differences in the translation task scores. To answer the first research question, our analyses were conducted with learners' subjective mental load ratings of difficulty in Phase 2 by running the Kruskal-Wallis Test. Another series of Kruskal-Wallis tests were performed to compare the levels of listening comprehension among the three groups in Phase 2 so as to answer the second research question. As for the third research question, Wilcoxon Signed Rank Tests were performed to compare the scores between the Phase 2 and Phase 3, which would help us understand the learners' learning to listen.

4 Results

4.1 Translation

The first analysis, involving a translation task, was conducted in order to check for differences of the three groups in terms of vocabulary learning in Phase 1. As shown in the descriptive scores (see Table 3), the three groups obtained a mean

Table 3: Descriptive scores of the translation.

	Listening + subtitles (n = 45)	Listening + script (n = 22)	Listening + only (n = 34)
Mean Score	9.16	9.18	9.26
Std. Deviation	0.85	1.09	0.93

of 9.20 out of 10, indicating that the ten target words were learned effectively. The scores on the translation task were submitted to the Kruskal-Wallis Test and the results showed no significant group differences ($H(2) = 0.71, p = 0.704, r = 0.046$). In this regard, it could be concluded that participants acquired the target words that were to appear in the subsequent listening texts in Phases 2 and 3. It also showed that the three groups were comparable in that they learned the target words to a similar extent.

4.2 Subjective rating of difficulty

The second analysis, which involved participants’ subjective mental load ratings, would allow the first research question to be answered. The descriptive data in Table 4 show that the subjective mental load ratings of difficulty were quite high in both Phase 2 and Phase 3, indicating that the two passages were fairly difficult for the participants in each of the three groups to understand. Further exploration of the data, by running independent samples Kruskal-Wallis Test and the Mann-Whitney Test, did not reveal any significant main effect of modes of presentation ($H(2) = 6.06, p = 0.048, r = 0.28$) in Phase 2. Yet seen from the descriptives, the participants in the Listening + subtitles group found the material relatively easier to comprehend than the Listening + script and Listening + only group; and listening with a full script was shown to be the most difficult mode for learners to understand the material.

4.3 Free recall

The third analysis, involving the test performance in the free recall task, would allow the second and third research questions to be answered. According to the descriptive data in Phase 2, the listening materials were difficult for the participants and a floor effect was identified in their test performance. All the three

Table 4: Subjective rating of difficulty and free recall scores.

		Phase 2–Coffee Trees			Phase 3–Roses		
		Listening + only (n = 34)	Listening + script (n = 22)	Listening + subtitles (n = 45)	Listening + only (n = 34)	Listening + script (n = 22)	Listening + subtitles (n = 45)
Difficulty Rating	M	7.79	8.00	7.28	7.97	8.27	7.98
	SD	0.68	1.11	1.35	0.76	1.03	0.94
Recall Task	M	1.24	1.41	1.16	0.82	1.05	0.84
	SD	1.20	1.14	1.38	1.27	1.52	1.11

groups achieved low scores with average less than 1.5 out of 12 units in the first test and less than 1.1 out of 11 units in the second one; no group differences in the free recall in Phase 2 were revealed by the Kruskal-Wallis Test ($H(2) = 1.55$, $p = 0.461$, $r = 0.09$). In Phase 3, again, no group differences in the free recall in Phase 3 were revealed ($H(2) = 0.286$, $p = 0.867$, $r = 0.08$), even though there is a slight non-significant tendency in the two listening with written text groups to get higher scores than the Listening + only group. A series of Wilcoxon Signed Rank Tests were then performed to compare the differences within each group between Phases 2 and 3. There was a tendency for all the groups to get lower scores; but again, the decline in the three groups did not reach statistical significance.

5 Discussion

In general, the results have shown that the listening materials were difficult for our participants, and that proficiency level of the students in this study and Diao et al. (2007)'s were quite different. While the students in the original study rated the difficulty of the text in Coffee trees' text with a mean that ranged from 4.04 to 4.84 depending on the group, those in our study rated the same text with a mean that ranged from 7.28 to 8 (in a nine-point scale where 9 indicates 'extremely difficult'). This difference in the proficiency levels of the two samples was unexpected since the two samples had had a similar number of years of English study and had a similar background in that the two were university students in China. We could not anticipate this as the original study did not provide any other indication of proficiency rather than the number of years of English study. In our study, the ratings of difficulty provided by our students were shown to be coherent with their self-reported proficiency of English with an average between 1.66 and 1.71 on a four-point scale where 1 indicated 'poor command'. Thus, our participants were at an unexpected lower level of English proficiency than those of the study we have replicated.

Past research has shown that there were differences in comprehension and perceived difficulty according to listening presentation modes. In those students, the level of proficiency of the students and the level of difficulty of the text was well-adjusted. However, in the present study there was an unexpected mismatch that stopped students from benefitting from any of the three presentation modes. So, when low proficient students encounter texts that they are not ready yet to process, they do not benefit from different presentation modes. In this study, the differences in ratings of difficulty were non-significant, but still it

can be seen from the descriptives that in Phase 2, those who listened with a full script found the material the most difficult to understand among the three modes of presentation. This could be due to that, for the low proficient learners, coping with reading a full script at the same time as listening to the oral text may have been cognitively demanding. The presence of the script may have produced a negative effect on the learners' perception of difficulty in listening hypothetically because the participants were low English level learners and their four English skills were not well developed. Thus, a task combining two language skills, in this case, listening to the material and reading an identical script, will possibly put learners in a difficult situation, dividing their attention and working memory resources to these two skills. According to CLT, redundant information will overburden the working memory and result in the redundancy effect (Sweller and Chandler 1994; Kalyuga et al. 2004; Sweller et al. 2011) especially among learners of low proficiency (Leslie et al. 2012). This finding accords with Chang's (2009) claim that listening while reading was more suitable for higher-level learners in that they had more advanced reading skills. In this study, the subtitle group rated the material easiest to comprehend, this can be because the listening + subtitles condition was a more guided, a strongly paced type of activity than the listening + script condition because students in the former condition only had access to the sentence that they could hear. Their attention was probably more easily focused, and this probably facilitated the breaking down of the sentences into coherent units and hypothetically contributed to a lower perception of difficulty in comprehension. As stated in Sweller (2017), implications of the CLT are especially relevant for beginner language learners and instructors should avoid learners' split-attention for different sources of information such as listening while reading with a full script, this also explicates, as seen in the results, why there was a trend for the listening only group to find the task easier than the listening + script group since they did not have to deal with any redundancy of input.

The effects of the written text on learners' comprehension were measured by the free recall task in Phase 2. The results revealed no significant differences among the three groups, which was also different from the results of Diao et al. (2007) and some other studies that claimed the positive effects of the written text (e.g. Moreno and Mayer 2002; Bird and Williams 2002). In the case of the present study, the learners could not benefit from the full presence of the written text, which can be explained by the fact that learners did not allocate their working memory resources in an efficient manner. Moreover, a floor effect was found in participants' free recall task scores, which indicated a mismatch between the difficulty of the listening passage and participants' proficiency level. For learners at a lower English level, when simultaneously presenting

the written text and spoken text, they are very likely to experience cognitive overload due to the redundancy effects and their limited working memory capacity. Hence no advantages to the written text were identified. The difficulty of the listening passage in the original study was matched with their participants' proficiency level, but it was not matched in the present study because participants were non-English major students and had a higher subjective mental load rating of difficulty. Accordingly, the results did not reveal a significant effect of either subtitles or a full script on low-proficient English learners' listening comprehension.

The final research question was answered by comparing the free task scores between Phase 2 and Phase 3 to test whether the presence of written text facilitates or interferes with participants' learning to listen. The statistical analyses did not report any significant differences between the two phases. All three modes did not show any positive or negative effects on learners' learning to listen as in the work of Diao et al. (2007). This can be explained that learners' low English proficiency was not matched with the text, and they were overwhelmed and found the listening passage too difficult for them. Fatigue factors may also lead to their poor performance on the second test. In our study, all three groups did not perform differently in the two phases, so we cannot arrive at a robust conclusion that whether written text is good for low-proficient learners' learning to listen or not, but there was a non-significant difference that those two groups with written text got slightly higher scores, and it could be, as illustrated by Charles and Trenkic (2015), that the written text in listening could enhance learners' speech segmentation abilities, thus further study could repeat the experiment with materials more attune to participants' level, which would produce more robust answers.

Therefore, seen from the findings, it should be noted that the difficulty of the listening passage should accord with learners' proficiency in order to show clearly the effect of the differing presentation modes. The findings also offer some methodological insights that a full and detailed description of participants, particularly the precise information of the proficiency level, should be included in SLA studies so that further replication research can be carried out. Otherwise, based on the present study, it is risky to carry out replication studies without such information.

6 Conclusion

To sum up, the present study explored the effects of the written text on learners' perceived listening difficulty, listening comprehension of spoken English and

their learning to listen by replicating the study of Diao et al. (2007) with low-proficient English learners. The results show that, the use of written text should be consistent and tuned with the learners' proficiency level, otherwise, learners especially those at a lower level, when presented with simultaneous written and aural input, will be overloaded and cannot make use of the written text for their listening comprehension and learning to listen.

From the results, some implications can be drawn for EFL instructors on teaching listening to low-proficient EFL learners. In general, they need to consider carefully their choice of teaching technique and exercises so as not to strain cognitive load and cause redundancy effects on learners. To be more specific, it should be noted that using written text in listening may not always be helpful, especially if the listening texts are difficult for students. They should be concerned with the difficulty of the listening passage as well as with learners' proficiency level. The written text may be beneficial to the improvement of listening comprehension and listening skills only when the difficulty of the passage is well adjusted to the level of proficiency of the learners, if learners find the listening passage difficult, it may make no difference in their listening comprehension and learning to listen.

Admittedly, the present study could be improved in some aspects. To illustrate, as a replication study, the present study was intended to use the original materials, but in retrospect, we found the mismatch between the difficulty of the passages and participants' level. Large standard deviation and a floor effect were also shown in the free recall scores, and this resulted in a non-normal distribution of the data. Further research should consider carefully the level of potential participants and choose level-appropriate materials with, for instance, slower speech speed or simpler syntax. Also, since the purpose of the free recall was to check their comprehension of the material, it could have been done in Chinese which would be easier for the participants and may trigger better results. A pilot study, conducted in advance, would also help to check the validity and reliability of the instruments and materials. What's more, in our study we used self-reported measure for English proficiency considering that it has produced favourable results in a majority of studies (Oscarson 1997, cited in Ellis and Barkhuizen 2005), however, the validity of the self-reported measure cannot be comparable to that of objective standardized proficiency test, thus future research should take this aspect into consideration. Moreover, as the acquisition of listening (and other) skills is a long-term process, firm conclusions cannot be fully achieved through a cross-sectional study, especially when looking at the development of a certain skill, more robust results can be obtained from a delayed post-test or a longitudinal design. Regarding the direction for future research, researchers could investigate the role of the difficulty of the text with learners of the same proficiency

level to compare whether listening with a full script has the same effect as listening with subtitles when listening to a text of different difficulty levels. Also, future research may investigate how the different listening modes have an effect on learners of different proficiency levels.

References

- Abbuhl, R. 2011. Why, when, and how to replicate research. *Research Methods in Second Language Acquisition: A Practical Guide* 13. 296.
- Baddeley, A. 1998. Recent developments in working memory. *Current Opinion in Neurobiology* 8(2). 234–238.
- Bandura, A. 1986. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.
- Bird, S. A. & J. N. Williams. 2002. The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics* 23(4). 509–533.
- Brown, D. & H. Lee. 2015. *Teaching by principles. An interactive approach to language pedagogy*. New York: Pearson.
- Chandler, P. & J. Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction* 8(4). 293–332.
- Chandler, P. & J. Sweller. 1996. Cognitive load while learning to use a computer program. *Applied Cognitive Psychology* 10(2). 151–170.
- Chang, A. C. S. 2009. Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System* 37(4). 652–663.
- Chang, A. C. S. 2016. Teaching L2 listening: In and outside the classroom. In W. A. Renandya & H. P. Widodo (eds.), *English language teaching today. Linking theory and practice*, 111–126. Switzerland: Springer.
- Chang, A. C. S. & S. Millett. 2014. The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal* 68(1). 31–40.
- Charles, T. J. & D. Trenkic. 2015. Speech segmentation in a second language: The role of bimodal input. In Y. Gambier, A. Caimi & C. Mariotti (eds.), *Subtitles and language learning: Principles, strategies and practical experiences*, 173–198. Bern: Peter Lang.
- Clark, J. M. & A. Paivio. 1991. Dual coding theory and education. *Educational Psychology Review* 3(3). 149–210.
- Diao, Y., P. Chandler & J. Sweller. 2007. The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology* 120(2). 237–261.
- Diao, Y. & J. Sweller. 2007. Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction* 17(1). 78–88.
- Ellis, R. & G. P. Barkhuizen. 2005. *Analysing learner language*. Oxford: Oxford University Press.
- Field, J. 2008. *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Geary, D. C. 2008. An evolutionarily informed education science. *Educational Psychologist* 43(4). 179–195.
- Graham, S. 2017. Research into practice: Listening strategies in an instructed classroom setting. *Language Teaching* 50(1). 107–119.

- Groot, A. D., F. Gobet & R. W. Jongman. 1996. Perception and memory in chess. *ICGA Journal* 19 (3). 183–185.
- Kalyuga, S., P. Chandler & J. Sweller. 2004. When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors* 46(3). 567–581.
- Kalyuga, S. & J. Sweller. 2014. The redundancy principle in multimedia learning. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning*, 2nd edn. 247–262 New York, NY.: Cambridge University Press.
- Kartal, G. & H. Simsek. 2017. The effects of audiobooks on EFL students' listening comprehension. *The Reading Matrix: An International Online Journal* 17(1). 112–123.
- Krashen, S. D. 1985. *The input hypothesis: Issues and implications*. New York: Addison-Wesley Longman Ltd.
- Leslie, K. C., R. Low, P. Jin & J. Sweller. 2012. Redundancy and expertise reversal effects when using educational technology to learn primary school science. *Educational Technology Research and Development* 60(1). 1–13.
- Mayer, R. & L. Fiorella. 2014. Principle for reducing extraneous processing in multimedia learning: Coherence, signalling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning*, 2nd edn. 279–315. New York, N.Y.: Cambridge University Press.
- Mayer, R. E., H. Lee & A. Peebles. 2014. Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology* 28(5). 653–660.
- Moreno, R. & R. E. Mayer. 2002. Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology* 94(1). 156.
- Morton, J. 1969. Interaction of information in word recognition. *Psychological Review* 76(2). 165.
- Paas, F., A. Renkl & J. Sweller. 2003. Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38(1). 1–4.
- Paas, F. & J. Sweller. 2012. An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review* 24(1). 27–45.
- Paivio, A. 1990. *Mental representations: A dual coding approach*. Oxford: Oxford University Press.
- Perez, M. M., W. Van Den Noortgate & P. Desmet. 2013. Captioned video for L2 listening and vocabulary learning 29: A meta-analysis. *System* 41(3). 720–739.
- Porte, G. K. 2015. Replication research in quantitative research. In J. D. Brown & C. Coombe (eds.), *The Cambridge guide to research in language teaching and learning*, 140–145. Cambridge: Cambridge University Press.
- Siegel, J. 2013. Exploring L2 listening instruction: Examinations of practice. *ELT Journal* 68(1). 22–30.
- Siyanova-Chanturia, A. & S. Webb. 2016. Teaching vocabulary in the EFL context. In W. A. Renandya & H. P. Widodo (eds.), *English language teaching today. Linking theory and practice*, 227–240. Switzerland: Springer.
- Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12(2). 257–285.
- Sweller, J. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4(4). 295–312.
- Sweller, J. 1999. Instructional design. In *Australian Educational Review*. 1501–1510
- Sweller, J. 2004. Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science* 32(1–2). 9–31.

- Sweller, J. 2009. Cognitive bases of human creativity. *Educational Psychology Review* 21(1). 11–19.
- Sweller, J. 2010. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review* 22(2). 123–138.
- Sweller, J. 2011. Cognitive load theory. *Psychology of Learning and Motivation* 55. 37–76.
- Sweller, J. 2016. Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition* 5(4). 360–367.
- Sweller, J. 2017. Cognitive load theory and teaching English as a second language to adult learners. *Contact Magazine* 43(1). 10–14.
- Sweller, J., P. Ayres & S. Kalyuga. 2011. *Cognitive load theory*. NY: Springer.
- Sweller, J. & P. Chandler. 1994. Why some material is difficult to learn. *Cognition and Instruction* 12(3). 185–233.
- Sweller, J., P. A. Kirschner & R. E. Clark. 2007. Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist* 42(2). 115–121.
- Tragant, E., C. Muñoz & N. Spada. 2016. Maximizing young learners' input: An intervention program. *Canadian Modern Language Review* 72(2). 234–257.
- Tricot, A. & J. Sweller. 2014. Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review* 26(2). 265–283.
- Vandergrift, L. 2007. Recent developments in second and foreign language listening comprehension research. *Language Teaching* 40(3). 191–210.
- Vandergrift, L. & C. C. Goh. 2012. *Teaching and learning second language listening: Metacognition in action*. Routledge.
- Vanderplank, R. 1988. The value of teletext sub-titles in language learning. *ELT Journal* 42(4). 272–281.
- Vanderplank, R. 2016. *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Springer.
- Winke, P., S. Gass & T. Syodorenko. 2010. The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology* 14(1). 65–86.
- Wolvin, A. D. & C. G. Coakley. 1985. *Listening*. 2460 Kerper Blvd., Dubuque, IA 52001: Wm. C. Brown Publishers.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/iral-2018-0350>).