



Comparison of input modes: L2 comprehension and cognitive load

Ceyhun Karabiyik *

Ufuk University, Faculty of Education, Department of Foreign Language Education, Ankara, Turkey ORCID: 0000-0003-1408-931X

Sezen Arslan

Bandirma Onyedi Eylul University, School of Foreign Languages, Balikesir, Turkey ORCID: 0000-0002-5784-5704

Nurdan Kavaklı Ulutaş

Izmir Demokrasi University, Faculty of Education, Department of Foreign Language Education, Izmir, Turkey ORCID: 0000-0001-9572-9491

Article history

Received:
04.04.2022

Received in revised form:
28.07.2022

Accepted:
23.08.2022

Key words:

Multimedia learning; L2 listening comprehension; cognitive load; instructional video

The current study investigated L2-based assumptions of the Cognitive Theory of Multimedia Learning and Cognitive Load Theory for the multimedia, modality, and redundancy principles. In this non-equivalent groups quasi-experimental design study, four groups of Turkish-speaking teacher trainees of the English language received a 12-minute non-paced lesson on harp seal pups that included English audio (audio group), English audio with video (video + audio group), English captions with video (video + text group), and English audio with video and captions (video + audio + text group). A comprehension test as well as measures for difficulty and effort rating were used to collect data. One-way between-groups analyses of variance (ANOVA) were conducted to determine the effects of different modes of presentation on participants' learning performance and cognitive load. Moreover, Tukey Honestly Significant Difference (HSD) tests were performed to determine the groups that differed from each other. The findings showed that the video + audio group performed better and reported less difficulty and effort expenditure in the foreign/second language (L2) listening comprehension task than the audio-only group. On the other hand, the video + text and video + audio groups did not differ with respect to comprehension, difficulty, and effort expenditure. Lastly, while the video + audio + text and video + audio groups performed equally well in the comprehension task, the video + audio + text group reported less difficulty and effort than the video + audio group. The results and possible venues for further research were discussed.

* Correspondency: ceyhun.karabiyik@ufuk.edu.tr

Introduction

It is reported half a decade ago that English is spoken by almost one-fifth of the whole world population (Lyons, 2017). This makes the value given to English as a medium for international interaction fair. For this reason, many countries teach English as a foreign language and make every endeavor to improve it. In so doing, the realities and needs of the current age are most recognized. To that end, the use of media and multimedia to provide input is common practice for many years in diverse learning settings (Anmarkrud, Andresen, & Braten, 2019) including face-to-face or online foreign/second language (L2) classes. Multimedia learning has many benefits such as providing flexible learning, active engagement, and interaction; and it caters to visual, aural, and verbal learning styles (Cairncross & Mannion, 2001), which resulted in unimodal and multimodal learning being an area of growing research interest. According to Lee and Mayer (2018), there are numerous studies on principles of instructional design using multimedia, yet most of these are in learners' native language. A growing number of studies were carried out in L2 settings focusing on the effect of media and multimedia on language skills development with comparisons among and between unimodal and multimodal presentations over the past two decades (see Zhang & Zou, 2021). Even though, the validity of the multimedia and redundancy principles for comprehension was tested in various L2 contexts before, including the Turkish L2 context; investigation of the validity of the modality principle in the Turkish L2 context is unique. Moreover, studies investigating the validity of multimedia, modality, and redundancy principles from a cognitive load (CL) perspective in L2 settings are scant with none identified in the Turkish L2 context. Therefore, the examination of the validity of the multimedia, redundancy, and modality principles in the Turkish L2 context is considered a valuable contribution to the existing line of literature.

Background

Cognitive Load Theory

Humans have a short-term and thus limited cognitive system, which is called 'working memory' (WM) (Sweller & Chandler, 1994). WM is mainly predicated on three significant properties: (1) dual channels that refer to separate storages for visual and verbal input, (2) limited capacity that denotes what quantity of each type of input can be processed at a time, and (3) active processing that entails active cognitive engagement for meaningful learning through choosing, arranging, and integrating the things (Mayer & Fiorella, 2014). Drawing on these inherent capacities, Cognitive Load Theory (CLT) explores to what extent humans can cope with the CL to start learning (Martin, 2014).

CLT extensively investigates how easily WM processes the information; thus, it is primarily focused on the impact of certain forms of load on WM (Sweller, van Merriënboer & Paas, 1998). These forms are (1) intrinsic CL which pertains to the inherent complex nature of the information or material to be mastered, (2) extraneous/ineffective CL which arises from irrelevant instructional procedures, and (3) germane/effective CL which springs from the way materials are presented (Paas, Renkl & Sweller, 2003). These three forms are tightly interwoven with each other, thereby impacting one another. For example, if learners partake in extraneous processing, the capacity to process other forms of the load is diminished at the same time by virtue of the limited capacity of WM (Lee & Mayer, 2018). As such, if the extraneous load does not overburden the WM capacity, learners can allocate germane resources for effective learning (Klepsch, Schmitz & Seufert, 2017). Therefore, CL is not only about the quantity of the items but also the items to be learned concurrently (Sweller &



Chandler, 1994).

To understand what type of load is addressed, it would be reasonable to investigate whether/to what extent the instructional task is cognitively difficult for the learners. There are several proposed measures but among them, the subjective measures (e.g., self-rating scale) are the most used references (Anmarkrud et al., 2019). It can be administered by asking the learners how much mental effort they have expended on the task and by exploring their perceived difficulty concerning the task (Paas, 1992). Given that the task demands much effort, learning can be hindered; therefore, an excessive CL must be avoided to maximize WM (de Jong, 2010). It would thus be useful to receive learners' self-ratings to inform and improve later instructional designs in L2 contexts.

Multimedia Learning (ML)

The theoretical underpinnings and implications derived from the CLT may be analyzed in a ML context because the learners need to recognize and process the information in varying modes and modalities (Brünken, Plass & Leutner, 2003). Therefore, it would be useful to clarify what modes and modalities are. Whereas the modes refer to the form of the presentation of the instruction such as words or illustrations, the latter denotes which processing channel (e.g., visual or auditory) is used to process the presented information (Mayer, 1997). Thus, multimedia instruction must consider brain and cognitive processes and recognize that WM has a limited capacity, thereby only processing some pictures in the visual channel and some sounds in the auditory channel (Mayer, 2009). In line with this, Mayer proposed 12 principles for multimedia design intending to organize and manage cognitive processing and to yield effective learning outcomes, labelled as (1) coherence, (2) image, (3) modality, (4) multimedia, (5) personalization, (6) pre-training, (7) redundancy, (8) segmenting, (9) signaling, (10) spatial contiguity, (11) temporal contiguity, and (12) voice (2009).

Since no ML design can promise perfect learning, learners' cognitive processes must be taken into consideration to understand how learning occurs (Ploetzner, Fillisch, Gewalt & Ruf, 2016). Principles for multimedia design are thus valuable as they are predicated on the operationalization of the WM and cognitive processing; and they comprehensively deal with how human learning improves with the true arrangement of visuals, verbal, and auditory information. Overall, they offer practical implications for the lecturers to design multimedia material for better learning.

Multimedia Learning: The Principles of Multimedia, Modality, Redundancy

As noted, Mayer (2009) listed 12 principles for ML. Of all, the multimedia principle constitutes the basic rationale for the rest of the principles (Butcher, 2014). Accordingly, human learning is more effective if the words are presented in integration with pictures rather than in isolation; thus, it allows verbal and visual representations to be held at the same time within WM. This would be beneficial for maximizing germane load while decreasing extraneous load.

Additionally, the modality and redundancy principles are particularly linked to the modes of multimedia thanks to the varying combinations of graphics, audio, and texts (Liu, Jang & Roy-Campbell, 2018). Otherwise, the modality principle asserts that humans can learn more efficiently from spoken words and pictures rather than pictures combined with printed words (Mayer, 2009). In other words, the integration of pictures with on-screen texts can overload

cognitive processing and hinder learning. In line with the basic principle of the human mind, that is visual and verbal input is stored in separate information processing channels (Mayer & Fiorella, 2014), presenting printed words and pictures simultaneously highly occupy each channel. Subsequently, this poses challenges against learning. Therefore, Mayer (2009) suggests presenting narrated texts and avoiding on-screen texts for better multimedia designs. Concerning the redundancy principle, Mayer (2009) asserts that humans can learn more effectively by means of pictures and spoken words than pictures with both spoken and printed words. He maintains if the words are introduced visually (i.e., on-screen texts), they put an extra CL on the visual information processing channel, thereby occupying other channels, and diminishing the capacity to process other modes of information. Therefore, he suggests removing the visually presented text to reduce extraneous processing. However, much has yet to be understood about the impeding effect of the redundant text. A related line of research (Ari, et.al, 2014) has contradictorily shown that redundant text may not always hinder learning but has the potential to improve the learning benefitted from the multimedia settings.

Input modes and L2 comprehension

Pursuant to the multimedia principle, learning is facilitated better when instruction is through a combination of words and pictures compared to when it is via words alone (Butcher, 2014; Mayer, 2009), even when the instruction is in L2, as picture presentation reduces the necessity for extraneous processing; thuswise releasing cognitive capacity for generative and essential processing required for comprehension (Lee & Mayer, 2015). Indeed, when instructed in L2, participants did learn better via multimodal presentation involving pictures and words in most studies (Chan, Lei & Lena, 2014; Lee & Mayer, 2015; Mayer, Lee & Peebles, 2014; Taşdemir, 2018; Yang, 2014). However, in a few cases, the audio-only groups were found to outscore groups instructed through multimedia presentations with words and pictures (Başal, Gülözer & Demir, 2015; Sarem & Marashi, 2020). Moreover, in some studies, no significant differences were revealed in this regard with respect to L2 comprehension (İnceçay & Koçoğlu, 2017; Matthew, 2020).

In consonance with the modality principle, learning is better facilitated when instruction is via spoken words and pictures than through the means of printed words and pictures (Low & Sweller, 2014; Mayer, 2009; Mayer & Pilegard, 2014). Nevertheless, when instructed in L2, learning is better facilitated through pictures and printed text than by the medium of pictures and spoken text (Lee & Mayer, 2018). This is because L2 instruction can cause heavy intrinsic CL necessitating heavy essential processing that can be eased via printed text (Lee & Mayer, 2018). In fact, when instructed in L2, a group instructed through spoken words and pictures was found to learn better than the group instructed through printed words and pictures (Syodorenko, 2010). In contrast to these findings, Lee and Mayer (2018) reported that participants receiving instruction via printed words and pictures outscored those instructed through spoken words and pictures. On the contrary, no significant differences were also revealed between a group receiving instruction in L2 via printed words and pictures and a group instructed in L2 through spoken words and pictures (Liu et al. 2018).

As per the redundancy principle, learning is better facilitated when instruction is through spoken words and pictures compared to when it is presented using pictures and corresponding spoken and printed words (Kalyuga & Sweller, 2014; Mayer, 2009; Mayer & Fiorella, 2014). However, when instructed in L2, learning is better facilitated via presentation consisting of pictures and corresponding spoken and printed words than presentation involving spoken

words and pictures (Sweller, Ayres & Kalyuga, 2011). The reason for this is that spoken text can impose a heavy extraneous load which can be eased with the addition of printed words (Sweller et al., 2011). Indeed, a considerable number of research in this respect concluded that when instructed in L2 presentation through pictures and corresponding spoken and printed words facilitated better learning than L2 presentation via spoken words and pictures (Aldera & Mohsen, 2013; Felek-Başaran, 2011; Goeman, De Grez & Deschacht, 2021; Hayati & Mohmedi, 2011; Kvitnes, 2013; Mirzaei, Meshgi, Akita & Kawahara, 2017; Lee & Mayer, 2018; Lin, Lee, Wang & Lin, 2016; Özgen 2008; Winke, Gass & Syodorenko, 2010). However, when instructed in L2, some groups instructed via spoken words and pictures learned better than groups instructed through pictures and corresponding spoken and printed words (Chen, Wang, Zou, Lin & Xie, 2019; İnceçay & Koçoğlu, 2017; Sarem & Marashi, 2020; Syodorenko, 2010). Moreover, there are also studies that revealed no differences between these groups instructed in L2 via these presentation modes (Chen et al., 2019; Chen et al., 2020; Hsu, Hwang, Chang & Chang, 2013; Kruger, Hefer & Matthew, 2014; Kruger & Steyn, 2014; Liu et al. 2018; Matthew, 2020; Mayer et al., 2014; Montero-Perez, Peters, Clarebout & Desmet, 2014; Şendurur, Doğusoy & Yondemir-Çalışkan, 2020; Wang & Tragant, 2019).

Input modes and cognitive load

In accord with the cognitive theory of multimedia learning (CTML) and CLT, instruction through words and pictures should be less difficult but more effortful for participants than the words-only mode of instruction even if it is presented in L2 (Lee & Mayer, 2015). Indeed, when instructed in L2, participants instructed via pictures and spoken words reported less difficulty and higher effort expenditure than those instructed through spoken presentation (Lee & Mayer, 2015). However, no differences were determined by a few studies with respect to either difficulty (Mayer et al. 2014) or effort expenditure (Matthew, 2020; Mayer et al., 2014) between these groups when the instruction is in L2.

According to the CTML and CLT, L2 instruction via printed words and pictures imposes less extraneous load than instruction through spoken words and pictures; therefore, it should be less difficult and require equivalent or lower levels of effort (Lee & Mayer, 2018). Indeed, when participants were instructed in L2 through printed words and pictures, they were found to experience less difficulty than those instructed via spoken words and pictures, but their levels of effort expenditure were equivalent (Lee & Mayer, 2018).

According to the CTML and CLT, L2 instruction through spoken and printed words with pictures imposes less extraneous processing than instruction via spoken words and pictures; therefore, it should be less difficult and require equivalent or lower levels of effort (Lee & Mayer, 2018). Indeed, when instructed in L2 participants instructed via spoken and printed words with pictures were found to experience less difficulty and expand equivalent or less effort compared to those instructed via spoken words and pictures (Kruger et al., 2014; Lee & Mayer, 2018; Lin et al., 2016; Mayer et al., 2014). But studies revealing equal difficulty levels are also present (Mayer et al., 2014; Şendurur et al., 2020; Wang & Tragant, 2019).

All in all, past research has yielded conflicting results with respect to the validity of the multimedia and redundancy principles with respect to both L2 comprehension and CL. Besides, whereas contradictory results are also evident for the validity of the modality principle with respect to L2 comprehension; it has been understudied from a CL perspective with only a single study (i.e., Lee & Mayer, 2018) to report to the best of our knowledge.

Moreover, the validity of the modality principle with respect to L2 comprehension and the validity of the multimedia, modality, and redundancy principles with respect to CL is unstudied in the Turkish L2 context to the best of our knowledge as well. For that reason, the current study is among the preliminary studies conducted in the Turkish L2 context in his regard. Motivated by these gaps in previous literature and Mayer's (2014) argument that it is more useful and promising to identify boundary conditions when certain principles do or do not apply, the current study aimed to investigate whether the multimedia, modality, and redundancy principles are applicable in an L2 context (i.e., Turkish) in which participants with an intermediate level of L2 (i.e., English) received a non-paced (i.e., natural pace of native speaker) lecture through different modes of media on a topic unfamiliar to them. Thus, the current study sought answers to the following research questions considering theoretical predictions:

- (1) Do participants learn better with video + audio than with audio when learning in L2?
- (2) Do participants experience less difficulty with video + audio than with audio when learning in L2?
- (3) Do participants exert more effort with video + audio than with audio when learning in L2?
- (4) Do participants learn better with video + text than with video + audio when learning in L2?
- (5) Do participants experience less difficulty with video + text than with video + audio when learning in L2?
- (6) Do participants exert less or equivalent effort with video + text than with video + audio when learning in L2?
- (7) Do participants learn better with video + audio + text than with video + audio when learning in L2?
- (8) Do participants experience less difficulty with video + audio + text than with video + audio when learning in L2?
- (9) Do participants exert less or equivalent effort with video + audio + text than with video + audio when learning in L2?

Methodology

Research Design

A non-equivalent groups quasi-experimental design was used in this study. This research design was adopted as the participants were assigned to different groups that differed from each other with respect to the mode of L2 input received. The inputs were identical in terms of content but differed with respect to their delivery format which were: (1) audio, (2) audio + video, (3) audio + text, and (4) video + audio + text. Concordantly, the L2 input delivery mode was the independent variable in this study, while (1) comprehension, (2) difficulty, and (3) effort were the dependent variables.

Participants

139 preservice English teachers (i.e., 44 females and 95 males) from a foundation university in Ankara, Turkey participated in the study. Their mean age was 19.93 (SD= 1.31). They were all native speakers of Turkish with at least a B2 (upper-intermediate) level of English proficiency as determined by the institutional English proficiency exam. In a



between-subjects design, participants served in the audio (n= 35), video + audio (n= 34), video + text (n= 33), and video + audio + text (n= 37) groups.

Materials

The instructional materials used in this study were the audio, video + audio, video + text, and the video + audio + text versions of a 12-minute non-paced video on harp seal pups (National Geographic, 2019). The materials were designed by National Geographic and are public on their YouTube channel. It was ensured that the material contained unknown content and was level-appropriate by consulting the participants and two English language teaching experts. The lesson contained 1198 words spoken in English by a British male presenter at a natural pace. The audio version described facts about harp seal pups, their relationship with the mother seals, and their race against time to learn basic survival skills amid the circumstances created by global warming. The video + audio version contained the same audio content and included a video showing the scenes in the Arctic that corresponded to the information presented in the audio. The video and audio ran synchronously, and both were constantly present during the lesson. For instance, when the audio described the kiss between the mother harp seal and the pup to recognize each other, it showed a scene of a nose-to-nose kiss between the mother harp seal and the pup; and when the audio described the ice surface breaking up, the video showed pieces of ice on the move. The video was not essential to understand the facts presented via the audio but added visual representations of the facts described. The content analysis of the audio-visual material showed that all aforesaid facts and events evident in the audio were portrayed in the video. Moreover, the video + text version included the video and captions whereas the video + audio + text version included the video, audio, and captions. Information from audio + text was redundant; in that, they communicated identical information (Kalyuga & Sweller, 2014).

Materials regarding the dependent measures were in English and comprised of a comprehension test (see Appendix A) and two cognitive load measures on perceived mental effort and task difficulty. The comprehension test was developed by the researchers. Also, an expert panel that included four academics in the field of English language teaching was consulted to ensure the validity of the items. The test was composed of 10 multiple-choice questions with four alternative answers; and it assessed both factual and inferential understanding. An example item is as follows: What can be said for the harp seals?, (A) Fat reserves can be found in the baby harp seals, (B) Females can give birth to more than one baby in one birth, (C) If baby harp seals are hungry, they can eat the dead pups, (D) The adult harp seals mate once each year The questions were scored as either incorrect (0 point) or correct (1 point); therefore, scores that could be received from the test ranged between 0 and 10. In terms of reliability, the Cronbach Alpha coefficient for the test across all participants (N= 139) was .52 which is considered sufficient for tests regarding material mastery (Chen, 2018).

The utilization of self-report rating measures to determine cognitive load is a common practice (e.g., Lee & Mayer, 2015, 2018; Schmeck, Opfermann, Van Gog, Paas, & Leutner, 2015), and often involves the assessment of perceived mental effort accompanied with a subjective measure of task difficulty (Brünken, Seufert, & Pass, 2010). Therefore, a mental effort measure (Paas, 1992) that asked participants to rate the mental effort they invested during learning, and a task difficulty measure (Kalyuga, Chandler, & Sweller, 1999) that is related to the rating of their perceived difficulty in understanding the lesson were utilized. The original mental effort measure developed by Paas (1992) is scored on a 9-point Likert scale

(from very, very low to very, very high) whereas the original task difficulty measure is scored on a 7-point Likert scale (from extremely easy to extremely difficult). To provide participants with a more coherent response format and allow for comparisons between the measures, the mental effort scale was adapted to a 7-point Likert format.

Procedure

After ethical clearance for the study was granted by the Social Sciences and Humanities Academic Research and Publication Ethics Committee of the institution at which the study took place, participants were informed about the study via announcements made during class hours. They were informed about the voluntary nature of participation, aims, and procedures of the study. The data collection was carried out with 139 voluntary participants who were randomly assigned to groups. First, explanations related to the procedures were made to the participants. Later, depending on the treatment group, the media or multimedia was presented, which was followed by the administration of task difficulty and mental effort measures, and the comprehension test. Participants were allotted five minutes to complete the rating scales and the comprehension test. Throughout the treatment and data collection process, which took about 20 minutes, participant rights to withdraw from the study were protected, and principles of human subject research were followed.

Data Analysis

Before conducting statistical procedures using SPSS version 25 to unearth group differences in comprehension, task difficulty rating, and mental effort, data were checked for accuracy, missing data, outliers, normality, and homogeneity of variance in line with the procedures and criteria asserted by Tabachnick and Fidell (2013). Data accuracy was checked via proofreading, whereas missing data were checked via frequencies. Data entry was accurate and there was no missing data. The data were screened for outliers using the Mahalanobis distance and it was determined that the data were free from outliers. Moreover, the normality of the data was evaluated via skewness and kurtosis, using .01 alpha level as the sample sizes in each group was small (p. 114). The analysis of the skewness and kurtosis values revealed normally distributed data. Lastly, the homogeneity of variance was assessed by the Fmax ratios in relation to the sample size ratios. As the sample size ratios were less than a ratio of 4 to 1, they were considered as relatively equal, in which case Fmax values as great as 10 are acceptable (p. 120). The Fmax ratios for the variables were $F_{\max(\text{comprehension})} = 1.28$, $F_{\max(\text{difficulty})} = 1.98$, and $F_{\max(\text{effort})} = 2.07$ respectively. As all Fmax ratios were well below the threshold level, they were all acceptable. After the assumption checks, a series of ANOVAs was performed to analyze whether the groups differed on the comprehension test, difficulty rating, and effort rating. Post Hoc comparisons were run by using Tukey HSD tests to find out which groups significantly differed from one another.

Results

Table 1 reports descriptive statistics (i.e., means and standard deviations) relevant to each treatment group on the comprehension test, difficulty rating, and effort rating. A series of One-way ANOVAs was conducted to determine whether the groups differed from each other with respect to their comprehension scores, difficulty ratings, and effort ratings. The results revealed that the groups differed with respect to the comprehension scores $F(3, 135) = 9.69$, $p = .00$, difficulty rating $F(3, 135) = 29.28$, $p = .00$, and effort rating $F(3, 135) = 33.31$, $p = .00$. In addition to statistically significant differences among groups with respect to comprehension scores, difficulty ratings, and effort ratings, the effect sizes computed using



eta squared also yielded large effects ($\eta^2_{\text{comprehension}} = .18$, $\eta^2_{\text{difficulty}} = .39$, $\eta^2_{\text{effort}} = .43$) (Cohen, 2016). This shows that the variances in comprehension scores, difficulty ratings, and effort ratings are indeed substantively related to the mode of presentation received by the participants in L2.

Table 1. Descriptive statistics on the comprehension test, difficulty rating, and effort rating for the four treatment groups

Group	Comprehension		Difficulty		Effort	
	M	SD	M	SD	M	SD
a. Audio	5.17 ^{bcd}	1.76	4.91 ^{bcd}	.61	5.34 ^{bcd}	.54
b. Video +audio	7.26 ^a	1.39	3.74 ^{ad}	1.14	4.26 ^{ad}	1.02
c. Video + text	6.33 ^a	1.79	3.70 ^{ad}	.81	4.12 ^{ad}	1.05
d. Video + audio + text	6.59 ^a	1.62	2.76 ^{abc}	1.21	3.08 ^{abc}	1.12

Note: Superscripts next to each mean value indicate the group(s) it significantly differs with at $p < .05$.

Do participants learn better with video + audio than with audio when learning in L2?

In accordance with the CTML and CLT, the video + audio group should score higher on a comprehension test than the audio group when learning in L2 (Lee & Mayer, 2015). Post-hoc comparisons using Tukey HSD tests (with $p = .00$) showed that the video + audio ($M = 7.26$, $SD = 1.39$) did indeed score significantly higher than the audio group ($M = 5.17$, $SD = 1.76$). The effect size ($d = 1.31$) was found to be greater than Cohen's (2016) guideline for a large effect ($d = .80$).

Do participants experience less difficulty with video + audio than with audio when learning in L2?

According to the CTML and CLT, the video + audio group should experience less difficulty than the audio group when learning in L2 (Lee & Mayer, 2015). Post-hoc comparisons using Tukey HSD tests (with $p = .00$) showed that the video + audio group ($M = 3.74$, $SD = 1.14$) did indeed experience less difficulty than the audio group ($M = 4.91$, $SD = .61$). The effect size ($d = 1.28$) was found to be greater than Cohen's (2016) guideline for a large effect ($d = .80$).

Do participants exert more effort with video + audio than with audio when learning in L2?

In agreement with the CTML and CLT, the video + audio group should exert more effort than the audio group when learning in L2 (Lee & Mayer, 2015). However, post-hoc comparisons using Tukey HSD tests (with $p = .00$) showed that the video + audio group ($M = 4.26$, $SD = 1.02$) exerted less effort than the audio group ($M = 5.34$, $SD = .54$). The effect size ($d = 1.32$) was found to be greater than Cohen's (2016) guideline for a large effect ($d = .80$).

Do participants learn better with video + text) than with video + audio when learning in L2?

In compliance with the CTML and CLT, the video + text group should do better on a comprehension test than the video + audio group when learning in L2 (Lee & Mayer, 2018). However, post-hoc comparisons using Tukey HSD tests (with $p = .10$) revealed a non-significant difference between the groups with respect to comprehension.

Do participants experience less difficulty with video + text than with video + audio when learning in L2?

In line with the CTML and CLT, the video + text group should less difficulty than the video + audio group when learning in L2 (Lee & Mayer, 2018). However, post-hoc comparisons using Tukey HSD tests (with $p = .99$) ascertained that the difference between the groups was non-significant as regards difficulty experienced.

Do participants exert less or equivalent effort with video + text than with video + audio when learning in L2?

In keeping with the CTML and CLT, the video + text group should exert less or equivalent effort with the video + audio group when learning in L2 (Lee & Mayer, 2018). Post-hoc comparisons using Tukey HSD tests (with $p = .93$) showed that there was no significance between the groups with respect to effort expended. In other words, the video + text group ($M = 4.12$, $SD = 1.05$) and the video + audio group ($M = 4.26$, $SD = 1.02$) did indeed expand equivalent effort.

Do participants learn better with video + audio + text than with video + audio when learning in L2?

As per the CTML and CLT, the video + audio + text group should do better on a comprehension test than the video + audio group when learning in L2 (Lee & Mayer, 2018). However, post-hoc comparisons using Tukey HSD tests (with $p = .32$) uncovered that the two groups did not differ with respect to comprehension.

Do participants experience less difficulty with video + audio + text than with video + audio when learning in L2?

As regards the CTML and CLT, the video + audio + text group should experience less difficulty than the video + audio group when learning in L2 (Lee & Mayer, 2018). Post-hoc comparisons using Tukey HSD tests (with $p = .00$) showed that the video + audio + text group ($M = 2.76$, $SD = 1.21$) did indeed experience less difficulty than the video + audio group ($M = 3.74$, $SD = 1.14$). The effect size ($d = .83$) was found to be greater than Cohen's (2016) guideline for a large effect ($d = .80$).

Do participants exert less or equivalent effort with video + audio + text than with video + audio when learning in L2?

In reference to the CTML and CLT, the video + audio + text group should exert less or equivalent effort with the video + audio group when learning in L2 (Lee & Mayer, 2018). Post-hoc comparisons using Tukey HSD tests (with $p = .00$) showed that the video + audio + text group ($M = 3.08$, $SD = 1.12$) did indeed expand less effort than the video + audio group ($M = 4.26$, $SD = 1.02$). The effect size ($d = 1.10$) was found to be greater than Cohen's (2016) guideline for a large effect ($d = .80$).



Discussion

This study researched the validity of the multimedia, modality, and redundancy principles from both ML and CL perspectives in an L2 context. The aim was to determine to what extent their theoretical assumptions including their boundary conditions could be extended when the material was non-paced in a L2 ML context.

The study revealed that participants learned better with pictures and spoken words (video + audio) than with spoken words (audio) when learning in L2. This result suggests that multimodal presentation in the form of video-aided audio was more facilitative in promoting learning in L2 than audio-only presentation. This finding contradicts past research most of which presented topics familiar to learners (Başal et al., 2015; İnceçay & Koçoğlu, 2017; Matthew, 2020; Sarem & Marashi, 2020). However, there are also past studies that support this finding (Chan et al., 2014; Lee & Mayer, 2015; Mayer et al., 2014; Taşdemir, 2018; Yang, 2014). In this regard, Plass and Jones's (2005) asserts that multimodal input in L2 with picture and spoken text enables learners to form verbal and visual mental models and establish links amongst them; and these in turn, enable them to retrieve learned information via two types of cues compared to only one when they with only spoken text (p. 480). Therefore, it can be concluded that when intermediate L2 learners are instructed in L2 on topics unfamiliar to them, learning can be better facilitated through a non-paced presentation involving pictures and spoken text than spoken text alone.

Moreover, participants experienced less difficulty with pictures and spoken words (video + audio) than with spoken words (audio) when learning in L2. Concordantly, this result suggests that multimodal presentation in the form of video-aided audio was less difficult when learning in L2 than audio-only presentation. This finding contradicts past research in which the participants' L2 proficiency was higher (Mayer et al., 2014). This finding supports previous research (Lee & Mayer, 2015) and the argument that the addition of redundant pictures to spoken text can reduce extraneous processing when learning in L2 (Lee & Mayer, 2015, p. 452). Therefore, it can be concluded that when intermediate L2 learners are instructed in L2, learning can be less difficult through a non-paced presentation involving pictures and spoken text than spoken text alone.

Participants also exerted less effort with pictures and spoken words (video + audio) than with spoken words (audio) when learning in L2. That is, this result suggests that multimodal presentation in the form of video-aided audio was less burdensome when learning in L2 than audio-only presentation. This finding contradicts previous research in which the participants were highly proficient in L2, received exam-oriented training in L2 listening comprehension, or were presented with topics familiar to them (Lee & Mayer, 2015; Matthew, 2020; Mayer et al., 2014;). This finding supports Lee and Mayer's (2015) contention that the addition of pictures to spoken text can reduce the cognitive effort necessary to ascertain word meaning. Therefore, it can be concluded that when intermediate L2 learners without listening-specific exam-oriented training are instructed in L2 on topics unfamiliar to them, learning can be a less effortful process through a non-paced presentation involving pictures and spoken text than spoken text alone.

On the other hand, participants learned equally with picture and written text (video + text) and with pictures and spoken text (video + audio) when learning in L2. Expressly, this result suggests that multimodal presentation in the form of video-aided audio and video-aided printed text were equally facilitative in promoting L2 learning. This finding contradicts past research in which participants were beginner-level L2 learners or received listening-specific

exam-oriented training (Lee & Mayer, 2018; Syodorenko, 2010). This finding supports previous research (Liu et al., 2018) and the claim that the modality principle may not apply when the verbal material is in long segments (Mayer, 2014, p. 12). Therefore, it can be concluded when intermediate L2 learners without listening-specific exam-oriented training are instructed in L2 on topics unfamiliar to them, learning can be equally facilitated through a non-paced presentation involving either pictures and written text or pictures and spoken text.

Moreover, participants experienced equivalent difficulty with pictures and written text (video + text) with pictures and spoken text (video + audio) when learning in L2. Put it differently, this result suggests that multimodal presentation in the form of video-aided audio and video-aided printed text were equally difficult. This finding contradicts past research in which participants received listening-specific exam-oriented training (Lee & Mayer, 2018). According to Lee and Mayer (2018), in a multimodal presentation in the form of pictures and spoken text, when the words in the spoken text are produced at a fast pace and are unfamiliar or difficult to encode for the learners, it may not be possible for them to fully encode each verbal word segment before the next one is narrated. As a result, this may create more essential CL for L2 learners when the transitory nature of the spoken text is considered where a word is gone once spoken (Lee and Mayer (2018). Likewise, the multimodal presentation involving pictures and onscreen text in which the topic is unfamiliar, non-paced written text (i.e., captions) may as well create essential CL equivalent to that of pictures and spoken text as written onscreen text is also transitory. Therefore, it can be concluded when intermediate L2 learners without listening-specific exam-oriented training are instructed in L2 on topics unfamiliar to them, learning can be equally difficult with a non-paced presentation involving either pictures and written text or pictures and spoken text.

Participants also exerted equivalent effort with pictures and written text (video + text) and with pictures and spoken text (video + audio) when learning in L2. Put another way, this result suggests that multimodal presentation in the form of video-aided audio and video-aided printed text necessitated an equal amount of effort. This finding is in line with previous research (i.e., Lee & Mayer, 2018). De Westelinck, Valcke, De Craene and Kirschner (2005) revealed that when learners have low prior knowledge of the subject material presented through multimedia, pictures that explain the information increases mental effort. Likewise, pictures accompanying the written and spoken texts may have led to equal amounts of effort expenditure on part of the students in both multimodal presentation groups due to their lack of familiarity with the subject matter as determined prior to the experiment. Therefore, it can be concluded that when L2 learners without prior knowledge of a subject matter are instructed in L2, learning can require equal amounts of mental effort expenditure with a non-paced presentation involving either pictures and written text or pictures and spoken text.

In addition to these, participants learned equally with pictures and corresponding spoken and printed text (video + audio + text) and with pictures and spoken text (video + audio) when learning in L2. Stated differently, this result suggests that multimodal presentation in the form of video-aided audio and printed text and video-aided audio were equally facilitative in promoting L2 learning. This finding contradicts past research that involves different participants (i.e., proficiency level, prior content knowledge, and training), materials (i.e., duration and non-redundant pictures), and procedures (i.e., in adequate duration between pre- and post-test). (Aldera & Mohsen, 2013; Chen et al, 2019; Felek-Başaran, 2011; Goeman et al., 2021; Hayati & Mohmedi, 2011; İnceçay & Koçoğlu, 2017; Kvitnes, 2013; Lee & Mayer, 2018; Lin et al., 2016; Mirzaei et al., 2017; Özgen 2008; Sarem & Marashi, 2020; Syodorenko, 2010; Winke et al., 2010). This finding supports previous research (Chen et al.,

2019; Chen et al., 2020; Hsu et al., 2013; Kruger et al., 2014; Kruger & Steyn, 2014; Liu et al. 2018; Matthew, 2020; Mayer et al., 2014; Montero-Perez et al., 2014; Şendurur et al., 2020; Wang & Tragant, 2019) and the explication that pictures with spoken text and pictures with both spoken and written text can both create essential overload when presented at a fast pace (Mayer & Pilegard, 2014, p.318). Therefore, it can be concluded when intermediate L2 learners without listening-specific exam-oriented training are instructed in L2 on topics unfamiliar to them, learning can be equally facilitated through a non-paced presentation involving either pictures and spoken text or pictures and corresponding spoken and printed text.

Moreover, participants experienced less difficulty with pictures and corresponding spoken and printed text (video + audio + text) than with pictures and spoken text (video + audio) when learning in L2. To be specific, this result suggests that multimodal presentation in the form of video-aided audio and printed text was less difficult when learning in L2 than video-aided audio. This finding contradicts past research in which participants were highly proficient in L2 or included participants whose prior knowledge of the content was not determined, or in which the material presented was very short (Kruger et al., 2014; Mayer et al., 2014; Şendurur et al., 2020; Wang & Tragant, 2019). This finding supports previous research (Lee & Mayer, 2018) and elucidation that the inclusion of written text identical to that of spoken text supplements spoken text that goes by quickly, which in turn reduces extraneous load and is reflected in lower difficulty rating (p. 652). Therefore, it can be concluded that when intermediate L2 learners are instructed in L2 on topics unfamiliar to them, learning can be less difficult through a non-paced presentation involving pictures with spoken and written text than pictures and spoken text.

Lastly, participants exerted less effort with pictures and corresponding spoken and printed text (video + audio + text) than with pictures and spoken text (video + audio) when learning in L2. In a sense, this result suggests that multimodal presentation in the form of video-aided audio and printed text was less burdensome when learning in L2 than video-aided audio. This finding contradicts past research in which participants' prior knowledge of the content was not determined (Kruger et al., 2014). This finding support previous research (Lee & Mayer, 2018; Lin et al., 2016; Mayer et al., 2014) and Lee and Mayer' (2018) explanation that the inclusion of written text identical to that of spoken text mitigates the heavy extraneous load that may be imposed by narration, which in turn is reflected in lower effort rating (p. 652). Therefore, it can be concluded that when intermediate L2 learners are instructed in L2 on topics unfamiliar to them, learning can be less effortful through a non-paced presentation involving pictures with spoken and written text than pictures and spoken text.

Conclusion

From a ML perspective, the results of the present study procured further evidence for the applicability of the multimedia principle and the disappearance of the modality and redundancy principles. These findings support Plass and Jones' (2005) argument that modality and redundancy principles may not apply in L2 learning contexts (p. 480). On the other hand, from a CL perspective, some findings contradicted theory-driven predictions with respect to the multimedia principle in terms of effort rating and modality principle in terms of difficulty rating. However, findings supported theoretical predictions regarding multimedia principle in terms of difficulty rating, modality principle in terms of effort rating, and redundancy principle in terms of both difficulty and effort ratings. All in all, this study

procures further indication for certain boundary conditions for three fundamental principles of ML in an L2 context.

To put it all in simple terms, the findings of this study imply that multimodal presentation of knowledge using L2 is more facilitative for learning than unimodal presentation and the increase in modes of presentation makes learning less difficult and effortful for learners. Based on these, as far as teaching in L2 is concerned course, lesson, and materials designers should opt for multimodal means of knowledge transmission when using media when and where possible. Moreover, they should try to increase the channels of input in multimedia presentation as much as possible based on the available technologies to ease learning.

Limitations and future directions

Firstly, this study was based on a short-term posttest; therefore, future studies can be conducted using a pretest-posttest design as well as multiple posttests to see the long-term effects of different input modes on learner comprehension. Secondly, this study was carried out with Turkish learners of English as a foreign language, so future studies can be undertaken within different L2 contexts. Thirdly, the material used in this study was a non-paced authentic video, thus, future studies can either replicate, or use paced materials to assess the effectiveness of input modes. Fourthly, this study was conducted with a group of students with at least a B2 level of proficiency; thus, future studies can either replicate or use groups of different language proficiency, or even compare groups with different levels of proficiency with respect to different input modes. Finally, this study included four input modes (audio, video + audio, video + text, video + audio + text); therefore, future studies can add mode groups or include groups with different modes of media.

Acknowledgement

This study was carried out by researchers who are members of Learning & Teaching Innovations Research & Development Group (LET-IN R&D).

References

- Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education*, 68, 60–75. <https://doi.org/10.1016/j.compedu.2013.04.018>
- Anmarkrud, Ø., Andresen, A., & Braten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist*, 54(2), 61-83. <https://doi.org/10.1080/00461520.2018.1554484>
- Ari, F., Flores, R., İnan, F. A., Cheon, J., Crooks, S. M., Paniukov, D., & Kurucay, M. (2014). The effects of verbally redundant information on student learning: An instance of reverse redundancy. *Computers & Education*, 76, 199-204. <https://doi.org/10.1016/j.compedu.2014.04.002>
- Başal, A., Gülözer, K., & Demir, İ. (2015). Use of video and audio texts in EFL listening test. *Journal of Education and Training Studies*, 3(6), 83-89. <https://doi.org/10.11114/jets.v3i6.1001>
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53-61. https://doi.org/10.1207/S15326985EP3801_7



- Brünken, R., Seufert, T., & Pass, F. (2010). Measuring cognitive load. In J. L. Pass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181–202). New York, NY: Cambridge University Press.
- Butcher, K. R. (2014). The multimedia principle. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 174–205). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.010>
- Cairncross, S., & Mannion, M. (2001). Interactive multimedia and learning: Realizing the benefits. *Innovations in Education and Teaching International*, 38(2), 156-164. <https://doi.org/10.1080/14703290110035428>
- Chan, C., Lei, W., & Lena, X. (2014). A study of video effects on English listening comprehension. *Studies in Literature and Language*, 8(2), 53-58. <https://doi.org/10.3968/4348>
- Chen, J. (2018). KR-20, in B. B. Frey, (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 932-936). London: Sage.
- Chen, M. P., Wang, L. C., Zou, D., Lin, S. Y., & Xie, H. (2019). Effects of caption and gender on junior high students' EFL learning from iMap-enhanced contextualized learning. *Computers & Education*, 140, 103602. <https://doi.org/10.1016/j.compedu.2019.103602>
- Chen, M. P., Wang, L. C., Zou, D., Lin, S. Y., Xie, H., & Tsai, C. C. (2020). Effects of captions and English proficiency on learning effectiveness, motivation and attitude in augmented-reality-enhanced theme-based contextualized EFL learning. *Computer Assisted Language Learning*, 1–31. <https://doi.org/10.1080/09588221.2019.1704787>
- Cohen, J. (2016). A power primer. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (4th ed., pp. (279–284)). Washington, DC: American Psychological Association. doi:10.1037/14805-018
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38, 105–134. <https://doi.org/10.1007/s11251-009-9110-0>
- De Westelinck, K., Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in social sciences: limitations of external graphical representations. *Computers and Human Behavior*, 21(4), 555-573. <https://doi.org/10.1016/j.chb.2004.10.030>
- Felek-Başaran, H. (2011). *The effects of captioning on EFL learners' listening comprehension: an investigative study with intermediate level university prep-school students* (Unpublished master's thesis). Eskişehir Anadolu University, Turkey.
- Goeman, K., De Grez, L., & Deschacht, N. (2021). Fostering Learning from Instructional Video: Effect of Subtitle Language on Content Recall. *Journal of Educational Multimedia and Hypermedia*, 30(3), 233-255.
- Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology*, 42(1), 181–192. <https://doi.org/10.1111/j.1467-8535.2009.01004.x>
- Hsu, C. K., Hwang, G. J., Chang, Y. T., & Chang, C. K. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Educational Technology & Society*, 16(1), 403–414.
- İnceçay, V., & Koçoğlu, Z. (2017). Investigating the effects of multimedia input modality on L2 listening skills of Turkish EFL learners. *Education and Information Technologies*, 22(3), 901–916. <https://doi.org/10.1007/s10639-016-9463-3>
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351–371.

- Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp. 247–262). Cambridge University Press.
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01997>
- Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49(1), 105–120.
- Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled Academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, 7(5), 1-15. <https://doi.org/10.16910/jemr.7.5.4>
- Kvitnes, I. E. N. (2011). *Subtitles in the second language classroom*. (Unpublished master's thesis). Norwegian University of Science and Technology, Norway.
- Lee, H., & Mayer, R. E. (2015). Visual aids to learning in a second language: Adding redundant video to an audio lecture. *Applied Cognitive Psychology*, 29(3), 445–454. <https://doi.org/10.1002/acp.3123>
- Lee, H., & Mayer, R. E. (2018). Fostering learning from instructional video in a second language. *Applied Cognitive Psychology*, 32(5), 648–654. <https://doi.org/10.1002/acp.3436>
- Lin, J. J., Lee, Y. H., Wang, D. Y., & Lin, S. S. (2016). Reading subtitles and taking notes while learning scientific materials in a multimedia environment: Cognitive load perspectives on EFL students. *Journal of Educational Technology & Society*, 19(4), 47–58.
- Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers & Education*, 127, 190-200. <https://doi.org/10.1016/j.compedu.2018.08.025>
- Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp. 227–246). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.012>
- Lyons, D. (2017, July 26). How many people speak English, and where is it spoken? *Babbel Magazine*. Retrieved from <https://bit.ly/3sg9RU7>
- Martin, S. (2014). Measuring cognitive load and cognition: Metrics for technology-enhanced learning. *Educational Research and Evaluation*, 20(7-8), 592-621. <https://doi.org/10.1080/13803611.2014.997140>
- Matthew, G. (2020). The effect of adding same-language subtitles to recorded lectures for non-native, English speakers in e-learning environments. *Research in Learning Technology*, 28. <https://doi.org/10.25304/rlt.v28.2340>
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1-19. https://doi.org/10.1207/s15326985ep3201_1
- Mayer, R. E. (Ed.). (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp.279-315). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.015>

- Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia Learning in a Second Language: A Cognitive Load Perspective. *Applied Cognitive Psychology*, 28, 653–660. <https://doi.org/10.1002/acp.3050>
- Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp. 316–344). Cambridge University Press.
- Mirzaei, M. S., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. *ReCALL*, 29(2), 178–199. <https://doi.org/10.1017/s0958344017000039>
- Montero-Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18(1), 118–141.
- National Geographic (2019, August 8). *The Harp Seal's Race Against Time - Ep. 5 | Wildlife: The Big Freeze* [Video]. <https://youtu.be/BF2TZq-ntRQ>
- Özgen, M. (2008). *The use of authentic captioned video as listening comprehension material in English language teaching*. (Unpublished master's thesis). Konya Selçuk University, Turkey.
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1) 1-4. https://doi.org/10.1207/S15326985EP3801_1
- Plass, J., & Jones, L. C. (2005). Multimedia Learning in Second Language Acquisition. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 467-488). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.030>
- Ploetzner, R., Fillisch, B., Gewalt, P.A., & Ruf, T. (2016.) The role of student-generated externalizations in strategic multimedia learning and how current (web-)technology fails to support learner engagement. *Interactive Learning Environments*, 24(7), 1610–1628, <https://doi.org/10.1080/10494820.2015.1041401>
- Sarem, S. N., & Marashi, H. (2020). The impact of input modality and sensory mode on Iranian intermediate EFL learners' listening comprehension. *Journal of Language and Translation*, 10(1), 133-153.
- Schmeck, A., Opfermann, M., Van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, 43(1), 93–114. <https://doi.org/10.1007/s11251-014-9328-3>
- Şendurur, E., Doğusoy, B., & Yoldemir-Çalışkan, N. (2020). Investigation of non-native learners' informal learning processes from cognitive-load theory perspective. *Interactive Learning Environments*, 28(1), 95–106. <https://doi.org/10.1080/10494820.2018.1517096>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185-233. https://doi.org/10.1207/s1532690xc1203_1
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review* 10, 251–296. <https://doi.org/10.1023/A:1022193728205>
- Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50–73.

- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson.
- Taşdemir, M. S. (2018). The effect of visual support on listening comprehension in Turkish elementary EFL learners. *Turkish Online Journal of English Language Teaching (TOJELT)*, 3(3), 111-127.
- Wang, X., & Tragant, E. (2019). The effect of written text on comprehension of spoken English as a foreign language: A replication study. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2018-0350>
- Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65–86.
- Yang, H. Y. (2014). Does multimedia support individual differences? – EFL learners' listening comprehension and cognitive load. *Australasian Journal of Educational Technology*, 30(6). <https://doi.org/10.14742/ajet.639>
- Zhang, R., & Zou, D. (2021). A state-of-the-art review of the modes and effectiveness of multimedia input for second and Foreign Language Learning. *Computer Assisted Language Learning*, 1–27. <https://doi.org/10.1080/09588221.2021.1896555>

Appendix A

1. It is pointed out that the Arctic Sea Ice ---.
 - A) covers a small portion of the ocean
 - B) has enlarged in three decades
 - C) is an indicator of global climate
 - D) broke up in 2016 and many pups drowned

2. Which of the following can be true about the harp seals?
 - A) Harp seals migrate from the south to the north
 - B) Baby harp seals can be fed by any mother harp seal
 - C) The baby harp seals must put on weight quickly
 - D) Swimming is not a must for the harp seals to survive

3. It is pointed out that the baby harp seals---.
 - A) spend time with their mother for a very long time
 - B) kiss their mothers if they are hungry
 - C) do not need help when they swim
 - D) have a yellow coat when they are newly born

4. What can be said for the harp seals?
 - A) Fat reserves can be found in the baby harp seals
 - B) Females can give birth to more than one baby in one birth
 - C) If baby harp seals are hungry, they can eat the dead pups
 - D) The adult harp seals mate once each year

5. It is true that---.
 - A) the Arctic Sea Ice causes global cooling
 - B) the Arctic Sea Ice is not broken in some years
 - C) the early break of Arctic Sea Ice is a threat to the pups

- D) the exact location of the Arctic Sea Ice is not known
6. It can be understood that---
- A) the pups are breastfed for more than ten days
 - B) the pups are the smartest animal in the world
 - C) the mother harp seal feeds the pup once a day
 - D) the harp seals are mammals
7. Which of the following is true about the harp seals?
- A) A slower pattern of growth is observed for the infants
 - B) The infants get protected in the sea by their mothers
 - C) The mature harp seals have a white coat
 - D) The infants can breathe underwater
8. Which one of the following is true?
- A) Saving polar bears is important to the human welfare
 - B) Shrinking the Arctic Sea Ice has only local effects
 - C) Land use has an impact on climate change
 - D) Arctic Sea Ice is important to human beings only
9. It is true that ---.
- A) the mother harp seals give birth when the ice starts to break
 - B) in recent years, it gets much more challenging to raise the pups
 - C) harp seal mothers cannot distinguish their pups
 - D) male harp seals help females to nurse their pups
10. It can be concluded that -----.
- A) there is no hope to harness clean energy
 - B) we are facing the deadly effects of global warming
 - C) saving polar bears is impossible
 - D) it is a turning point for deforestation