OPEN ACCESS

CONTEMPORARY EDUCATIONAL TECHNOLOGY

ISSN: 1309-517X (Online) 2020, 12(1), ep262, https://doi.org/10.30935/cedtech/7628

Research Article

Investigating the Use of Text Positions on Videos: An Eye Movement Study

Hamza Polat
Erzincan Binali Yıldırım University, Turkey
ORCID: 0000-0002-9646-7507

Received: 23 Nov 2019 Accepted: 21 Jan 2020

Abstract

Videos have become an indispensable part of both online and blended learning environments. However, the design of such videos requires careful consideration of multimedia learning principles to reduce the cognitive load during the instruction. In this regard, the purpose of this study is to investigate the effect of text-positions presented at two videos on eye-fixation duration and remembering. An experimental research with one-shot case study design was employed to meet this purpose. Two videos about financial issues were selected from a public TV channel archive: one of them included on-screen texts located at the bottom, and the other included informative texts located on the right side of the screen. A total of 61 students first watched these videos by interacting with an eye-tracking device in a human-computer interaction lab and then completed a retention test. The results indicated a significant positive correlation between total eye fixation duration and retention test scores. Additionally, the fixation duration of the participants was higher when the texts were presented on the right side of the videos than that when texts were presented at the bottom. The total fixation durations for longer text were higher than those for shorter text.

Keywords: text position, multimedia learning, eye-movements, retention, eye-fixation duration

INTRODUCTION

Videos are potentially useful media in learning (Gaudin & Chaliès, 2015) that appeal to different senses of learners in an educational setting. They may provide both verbal and non-verbal stimulus, which makes them more suitable for human cognition to process information (Paivio, 1986). They allow students to experience the learning content at their own learning pace (Leo & Puzio, 2016) and provide a flexible learning environment in which learners pause, replay and rewind the learning content (Howard, Meehan, & Parnell, 2018). Along with the pedagogical functions these videos have and the rapid developments in information technologies, their popularity has increased in education and training in recent years (Bétrancourt & Benetos, 2018; de Koning, Hoogerheide, & Boucheix, 2018). They have been used as a multimedia learning tool in traditional (Nadelson, Scaggs, Sheffield, & McDougal, 2015) blended (González-Gómez, Jeong, Airado Rodríguez, & Cañada-Cañada, 2016) and online (Cooper & Higgins, 2015; D. Zhang, Zhou, Briggs, & Nunamaker, 2006) learning environments.

As a form of multimedia material, videos include both texts and visuals simultaneously (Mayer & Moreno, 2010). The design-related issues of these materials consist of the psychological aspect of message design (Zhang, 1996) to which educators need to pay more attention to reduce the extraneous cognitive load and to facilitate the essential processing during the learning (Mayer, 2001). In this regard, both the cognitive theory of multimedia learning (CTML) and cognitive load theory (CLT) provide several strategies to make them more suitable for human cognitive architecture (Fiorella & Mayer, 2018; Sweller, van Merriënboer, & Paas, 2019).

Several design strategies have been employed by a variety of researchers in recent years to understand whether they improve learning outcomes in multimedia learning environments. While some of the studies (Biard, Cojean, & Jamet, 2018; Boucheix, Gauthier, Fontaine, & Jaffeux, 2018; Colliot & Jamet, 2018) have reported positive results, others (Hoogerheide, van Wermeskerken, van Nassau, & van Gog, 2018; Merkt, Ballmann, Felfeli, & Schwan, 2018; van Wermeskerken, Ravensbergen, & van Gog, 2018) have not found an effect. Biard et al. (2018), for example, have reported the segmenting technique effective in managing essential processing during the learning, implicating that complex video lessons should be segmented into predetermined meaningful units. Likewise, it has been found that different camera viewpoints on videos may enhance students learning performance (Boucheix et al., 2018), suggesting that the camera perspective can be switched during the instruction depending on the demands of the learning task. On the other hand, Hoogerheide et al. (2018) have found that students perceive the same-gender models on the videos as more similar to them. However, they have not provided clear evidence about the influence of this similarity on learning performance and self-efficiency. Similarly, it has been reported that the presence of the instructors' image on the videos may cause extraneous processing, and thereby may draw learners away from the learning material (van Wermeskerken et al., 2018), but improves the learner's engagement and motivation (Colliot & Jamet, 2018). Given the related researches in the field, the research trend in multimedia learning can be categorized into three groups: (1) extending the scope of existing multimedia design principles, (2) providing new design principles, and (3) integrating individual differences to the designing process of multimedia materials (de Koning et al., 2018). From this point of view, there is still a need for evidence-based principles to understand in which condition and for which content instructional videos with multimedia principles do and do not enhance educational outcomes.

How to present informative-texts on videos is also a design-related issue of multimedia materials. Any empirical evidence about this issue may extend our understanding of the spatial contiguity principle and may give insights into new design strategies for information-rich video lessons. From this point of view, while preparing and presenting the informative texts on videos, the dual-channel and the limited capacity assumptions of the multimedia learning need to be considered. The dual-channel assumption takes its origin in the dual coding theory (Paivio, 1986) and insists that humans possess two separate channels for visual and auditory information. The capacity of each channel is limited in memory. Therefore, only a limited amount of information takes place at one time in the channels (Mayer & Moreno, 2003). When the cognitive demands of the learning task exceed the limited capacity in one or both channels, the learner becomes cognitive overload (Sweller et al., 2019), a situation that influences learning adversely (Mayer, 2001). In this regard, the position and the length of the informative-texts may influence extraneous cognitive and consequently increase or decrease the working memory resources that are essential to facilitate learning.

Mayer and Moreno (2003) provided several overload scenarios for multimedia learning environments. One of these is the overload caused by the confusing design of the material. They suggested two strategies to deal with this problem: eliminating the redundancy and aligning. The redundancy effect can be eliminated by avoiding the use of the on-screen texts with their narrations. The spatial contiguity effect, on the other hand, can be ensured by aligning on-screen texts near to the corresponding visuals. The literature provides empirical evidence for the contiguity effect that students perform better learning outcomes when texts are integrated to the visuals than texts and visuals presented separately (Bétrancourt & Bisseret, 1998; Ginns, 2006; Johnson & Mayer, 2012; Mayer, 1997). It might be beneficial to note that in an integrated design, these two sources of information (i.e., visuals and texts) should be unintelligible in isolation (Sweller, Ayres, & Kalyuga, 2011). In other words, if a visual provides all the information that is necessary to learn, there is no need to integrate texts to the visuals. However, there are also several cases where the texts are presented as an additional information source on videos, which are comprehensible in isolation. How viewers inspect these materials and how they process the texts presented in different positions of the screen may be valuable for educators to design video lessons in a way that may reduce the cognitive load. In this sense, the results of the present study may include insights into the design of these video materials in online-learning environments.

Videos in Education

In the 2000s, researchers in the field have asked the question that ‘do dynamic visualizations improve learning’ (Bétrancourt & Benetos, 2018). Since then, a variety of studies have investigated the use of videos in different settings in terms of different educational outcomes. Some of them have indicated that instructional videos are an effective tool for enhancing students’ learning performance (Love, Hodge, Grandgenett, & Swift, 2014; Nagy, 2018; Tang et al., 2017; Traphagan, Kucsera, & Kishi, 2010; van der Meij, 2017), satisfaction (Nagy, 2018), motivation (Tang et al., 2017; van der Meij & van der Meij, 2013), and task completion (van der Meij, 2017). On the other hand, the current literature has questioned ‘why and when do these materials improve learning’ and ‘for whom these materials are more effective’(Bétrancourt & Benetos, 2018). In the commentary by Bétrancourt & Benetos (2018), three levels of explanations, addressing representational, cognitive and perceptual, and instructional approaches, have been provided to answer these questions.

First, the representational approach refers to “identifying the semiotic and communicational properties of video compared to other types of representational formats” (Bétrancourt & Benetos, 2018, p. 472). It is assumed that videos support learning performance once they have been prepared by following the cognitive demands of the learner. It is also asserted that dynamic visualizations are more effective than static visualizations in conveying more complex information (Tversky, Morrison, & Betrancourt, 2002) and for dynamic contents that include continuous events changing over time. Second, cognitive and perceptual approaches focus on the design aspects of dynamic visualizations that may help learners to allocate cognitive resources to the essential part of the learning material. This approach is common in multimedia literature because of the prevalence of cognitive theories in the field. In this regard, the related literature has been growing by providing novel design principles or extending “traditional” ones (de Koning et al., 2018). Finally, videos have been used in different instructional purposes: (1) engaging learners with learning resources, (2) demonstrating how to perform a procedure, (3) giving opportunity to observe experts in real-life situations, and (4) providing a learning output that may have potential to make learners more active in learning process (Bétrancourt & Benetos, 2018).

Eye-Movements during a Cognitive Processing Tasks

The number of studies regarding eye-movements has increased in recent years, thanks to the availability of advanced video-based eye-tracking systems (Rayner, 2009; Schütz, Braun, & Gegenfurtner, 2011). These systems have been employed to test the effectiveness of multimedia learning principles during instruction (Mayer, 2010) and may be used to gain insight about the cognitive load (Huh, Kim, & Jo, 2019; Sweller et al., 2011). Modern eye-trackers allow researchers to analyze the areas that frequently interest users.

The information gathered from these eye-tracking devices can be used for at least three purposes depending on the research goal: (1) how certain multimedia-effects (e.g., redundancy, spatial contiguity, etc.) occur, (2) making improvements on the design of the multimedia materials, and (3) providing access to cognitive process of learners during the interaction with the material (van Gog & Scheiter, 2010). The present study focuses on the second purpose by aiming to understand how to present a large amount of text information based on the eye-movement patterns of learners while processing predetermined text positions on financial videos.

Researchers in applied psychology are highly interested in two core concepts about eye movements: fixations and saccades (Rayner, 2009). Fixation refers to the time interval when eyes remain still on something to acquire new information (Henderson, 2017; Rayner, 2009). This human gaze control is operated by the visual properties of a particular scene or the cognitive demands of a particular task (Henderson, 2007). Saccades, on the other hand, are the fast eye-movements that reorient the fixations from one point of interest to another during a task (Henderson, 2017; Noton & Stark, 1971; Rayner, 1998). These eye movements can be measured by eye-tracking systems and may tell more about how visual perception occurs during cognitive processing tasks (Noton & Stark, 1971; Reichle, Pollatsek, Fisher, & Rayner, 1998), including reading (Drieghe, Rayner, & Pollatsek, 2005; Rayner, Slattery, Drieghe, & Liversedge, 2011), visual search, scene perception (Rayner, 2009). Results from early studies demonstrated a consistent connection between saccades and attention (Deubel, 2008; Gersch, Kowler, Schnitzer, & Dosher, 2009; Hoffman & Subramaniam, 1995), fixation and attention (Henderson, 2007), and fixation and depth of learning (She & Chen, 2009). It was also shown that there is a correlation between saccades and acquired information detected from a scene to which the learner makes a saccade (Hoffman & Subramaniam, 1995).

Gaze control can be described as the process of directing fixation to a particular region of a scene during a perceptual, cognitive, and behavioral task (Henderson, 2003). Several factors may affect this behavior, including salience, recognition objects, plans, and values (Schütz et al., 2011). First, the stimulus salience has a moderated effect on guiding eye movements (Schütz et al., 2011). In this regard, visual aspects of a scene have a critical role in driving initial fixation placement (Henderson, Weeks, & Hollingworth, 1999). In simple terms, a colorful or a bright design may attract the eye movements (Henderson, 2007). For example, Ozcelik, Karakus, Kursun, and Cagiltay (2009) have reported that color-coded materials draw more learner attention than those of monochrome materials.

Second, learners tend to recognize objects rather than their features in general due to the natural assumption that saccadic target selection is driven by objects (Schütz et al., 2011). During a well-learned task, for example, viewers first scan the location of relevant objects in the setting before completing the task (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Henderson, 2017) and eye-fixations are oriented towards the objects being manipulated in this setting (Land, Mennie, & Rusted, 1999).

Third, eye tracking and neural recording studies have shown the influence of tasks on eye movements (Hayhoe & Ballard, 2005). Learners are persuaded by several goals to accomplish these tasks. In doing so, eye movements are mediated by relevant task demands (Schütz et al., 2011). Even if some part of the material is visually unnoticeable, viewers may direct their fixation location towards the task-relevant regions of a scene (Henderson, 2007). For example, Land et al., (1999) investigated the eye movement patterns during a primary everyday task. They reordered eye movements of subjects while they were making tea in a kitchen. The overall goal, making tea, consisted of a group of sub-goals such as filling the kettle, warming water to the teapot, and so forth. These lower-level visual and motor events were described as “object-related actions.” What researchers found was that gaze control behavior of the subjects was closely related to the objects (e.g., kettle, taps, etc.) that were related to these actions.

Finally, although it has been disregarded so far, values play an important role in eye movements (Schütz et al., 2011). Physical actions of people like moving the hand may cause several positive or negative consequences in the environment. These consequences may lead to further actions of people in a given situation. Particularly, tasks, including reward expectations, play an important role in guiding eye movements (Hayhoe & Ballard, 2005). Similarly, the term value refers to the expected reward or gain when selecting targets or deciding to make eye movements. In this regard, it has been shown that saccadic-eye movements are closely related to the relative value of actions (Platt & Glimcher, 1999). Furthermore, the level of interest (Catrysse, Gijbels, & Donche, 2018) and experience in the domain specific knowledge may also affect the eye-movement patterns during a scene perception or a reading task.

Text Positions at Multimedia Materials

The organization of the information on the screen has a crucial role in effective display design (Austin, 2009). As a display element, informative texts are attached to the visuals to facilitate learners’ perception of the message. There is empirical evidence that the integration of texts to visuals improves learning (Bétrancourt & Bisseret, 1998). Likewise, such an argument led to the contiguity principle in multimedia learning (Mayer, 2001). This principle asserts that the corresponding texts and visuals should be presumably presented with a small distance (Johnson & Mayer, 2012; Mayer, 2001; Moreno & Mayer, 1999). In this regard, Ginns (2006) conducted a meta-analysis by synthesizing the results of 37 studies and reported large mean effect size (d =.72) for spatial contiguity principle, providing some evidence in favor of using an integrated design (i.e., texts are located near to the corresponding visuals) in multimedia learning.

Figure 1. Text positions located at the videos

An integrated screen design can be in two forms in digital contexts: static texts on static visualizations (e.g., graphics) and fleeting texts on dynamic visualizations (e.g., videos). It is relatively easier to apply spatial contiguity principles on static visualizations (Schmidt-Weigand, Kohnert, & Glowalla, 2010) by locating semantically relevant texts near to the corresponding visuals. However, reading a changing text over time as well as scanning dynamic visuals and listening audios may cause cognitive load during the information processing (Kruger & Steyn, 2014).

Mayer, Steinhoff, Bower, and Mars (1995) investigated how an integrated design could be applied to illustrations of science textbooks. Their study provided insight into the textbook design with annotated illustrations having both informative texts and corresponding pictures. The contiguity principle was also suggested for instructional animations to promote learning (Mayer & Anderson, 1992; Mayer & Moreno, 2002; Mayer & Sims, 1994). On-screen texts should be close to their related animations rather than being presented far apart (Mayer & Moreno, 2002). However, setting the text layout on the screen is a complicated issue in interactive environments like 3D animations (Chigona, Sonnet, Ritter, & Strothotte, 2003). Considering the possibility that a standard text may block the appearance of a 3D object whose position changes in time, Chigona et al. (2003) suggested a shadow model in which texts are presented behind objects. Additionally, it has been shown that instructional videos having informative texts at the bottom of the screen (subtitle) are more effective in teaching vocabulary concepts than pictures with texts (Al-Seghayer, 2001). However, there is still no precise design principle suggesting the best text positions for instructional videos to which students allocate more cognitive resources.

Purpose of the Study

Contrary to the static visualizations, a video allows for multiple presentations of information. In other words, the message can be supported with images, graphs, texts, and sounds in a way that complements each other. Examining the eye-movements of the learners while they are interacting with these videos may provide valuable data about how they perceive the dynamic on-screen texts and where they look at during the interaction. Therefore, the purpose of this study is to investigate the use of on-screen texts presented at different positions of a finance video (see **Figure 1**) in terms of retention and eye-fixation duration. The study may also provide insight into how learners process the short and long texts (Drieghe et al., 2005; Rayner et al., 2011) at different segmentation of a video based on eye-movement patterns. Furthermore, the study may contribute to extend our understanding of integrated screen design in an information-reached multimedia learning environment. In this manner, we expect to propose suggestions for those who are dealing with an intensive amount of text on videos. Consequently, the following research questions guided to the study:

- Is there a correlation between total fixation duration during viewing a specific text position and the retention score gathered from the fixed text position?

- Is there a mean difference between the total eye-fixation durations of the participants at the bottom of the screen and on the right side of the screen?

- Is there a mean difference between the retention mean scores obtained from the bottom and right text positions presented on each video?

METHOD

Participants

The sample was taken from both graduate (*n*= 13) and undergraduate (*n*= 51) students from the department of educational technology at a state university in Turkey. While selecting the participants, the following criteria were addressed; (i) students should be native Turkish speaker since the material was prepared in Turkish, (ii) low experience in finance, and (iii) volunteered to continue participation in the study.

A questionnaire was developed to determine participants' experience in financial issues. A total of 77 students completed the test. Consequently, 64 students (33 males, 31 females) having low experience in financial issues were selected based on the responses gathered from the questionnaire. Students whose experience score higher than six did not involve in the study, which was done to control for the biasing effects of the experience as the split attention effect is smaller for high-experience students than those of low-experience students (Moreno & Mayer, 1999). Participants' age ranged from 20 to 28 years (*M*= 22.13, *SD*= 2.50).

Before performing the related analysis, the data were cleaned to remove participants whose weighted gaze sampling frequency below 90%. This was done because a lower percentage implies that the participant looked away from the monitor, which was not the desired situation in this study. Therefore, those three cases whose sampling frequency was 46%, 73%, and 81% were removed from the dataset to ensure the data quality. Consequently, the analysis was conducted with 61 cases whose weighted gaze sampling frequency ranged from 94% to 98% (*M*= 95.48, *SD*= 1.01).

Procedure and Setting

A one-shot case experimental design was employed in the present study. A pilot study was conducted in advance with a PhD candidate who had experience in eye movement studies. The main purpose of the pilot study was to foresee any possible problems we might encounter during the real experiment process and to provide adequate solutions for these problems in advance. During the pilot study process, what was expected from the participant was clearly explained, and the participant was allowed to think aloud. The participant interacted with the media by using the eye-tracker in the human-computer interaction (HCI) lab. After completing two instructional videos, the retention test was applied. Based on the opinions of the expert, and observations of the researchers, the actual design of the implementation process was finalized.

Three days before the actual implementation, all of the participants (*n*= 77) filled the prior knowledge test. A total of 65 students whose total prior knowledge score below seven was invited to the study. Because one of the selected students gave up to participate in the study on the day of actual implementation, the study was conducted with a total of 64 students. First, video materials were uploaded to the eye-tracker device within a special software. Then, all participants were called into the HCI laboratory one by one. The experimenter informed each participant that he or she would be interacting with two videos about financial issues, that his or her eye-movement patters would be recorded, and he or she would be receiving a retention test about the issues presented at the videos. They were also informed that the purpose of the experiment was to understand how they process the video materials (without evoking their covert attention towards the text positions). Before starting the eye-tracker calibration process, they signed the consent form.

Each participant was seated in a comfortable position in front of the eye tracker monitor, in a distance that was approximately 60 cm away from the monitor. The 16-point calibration process was initiated for each participant at the beginning of each session. After the calibration of the eye-tracker, the participant kindly informed to watch the first video material without moving his or her head. As soon as the first video display was completed, the second video was presented. The order of the videos was counterbalanced across participants. In other words, the videos were not played in the same order for every participant to minimize serial-position effects. After both of the videos were watched, each participant received approximately two minutes rest before completing the retention test.

Apparatus

A Tobii 1750 binocular eye-tracking system was used to monitor and record the eye movements of participants during the interaction with video materials. The system hardware is integrated into a 17-inch TFT display with a resolution of 1280–1024 pixels. The sampling rate was 50 Hz. The spatial resolution of the system was less than 0.5°. Four text positions illustrated in **Figure 1** were determined as “areas of interest” (AOI) on the system. To determine the eye-fixation duration on these AIOs, the criteria that the eyes remained fixed more than 100 milliseconds on a 50-pixel area was used (Duchowski, 2003). The Tobii Studio 2.2.8 software was used to calibrate the eyes of participants with the eye-tracker system.

Data Collection

The data directly related to the research questions were collected from three sources: prior knowledge and retention test, and eye-tracker records. The prior knowledge test was administered in the class. The total experience score was obtained from the prior knowledge test. Additionally, the retention test was administrated in the HCI laboratory, and each participant was given adequate time to complete the test. The correct answers of questions were counted for each participant to derive four retention scores addressing four text regions.

The eye-tracker, on the other hand, provided two types of data for inferential and descriptive statistics. For the descriptive part of data collection, the users’ eye movements were analyzed based on the gaze-plots. By this way, it is intended to figure out where the participants were focusing on while watching the media. For the inferential part of data collection, area of interests (AOIs) were selected on each video, which are the text regions illustrated in **Figure 1**. This method allowed us to collect all the eye-fixation data and only the data directly related to the presentation of the informative texts. In the end, total fixation data (in seconds) were obtained for each of the defined text positions.

Materials

Two types of materials were used in the study: videos and paper-based materials. First, two videos were prepared to collect information about the gaze control behavior of students. The videos were about financial issues in the stock market. The videos displayed reports of financial news by an expert in the economy, publicly broadcast on a national TV channel in Turkey. Both of the materials focused on similar issues about the topic. The duration of each video was approximately 3 minutes. They were simple and straightforward to present the main message.

Two text positions (layouts) were identified for each video by considering the text positions commonly used in daily life: the bottom and right side of the screen (see **Figure 1**). Each video also included areas of short and long text. Short-text areas have been prepared in a way that includes brief information about the daily foreign exchange rates in the market. Likewise, long-text regions have been designed to present important global business and market news in a sentence format. Consequently, the first video included short and long text regions at the bottom; similarly, the other one also included short and long text regions on the right side.

Attention allocation is generally affected by eye-movement patterns such that more eye-fixation directed towards a particular part of the screen reflects a considerable attention engagement on that part (Ferber & Murray, 2005). As noted by Schütz et al. (2011), the salience of the material is one of the factors that may influence the visual perception. In this regard, the following strategies have been employed to control the attention engagement of students that may be derived from external factors related to the visual aspects of the videos:

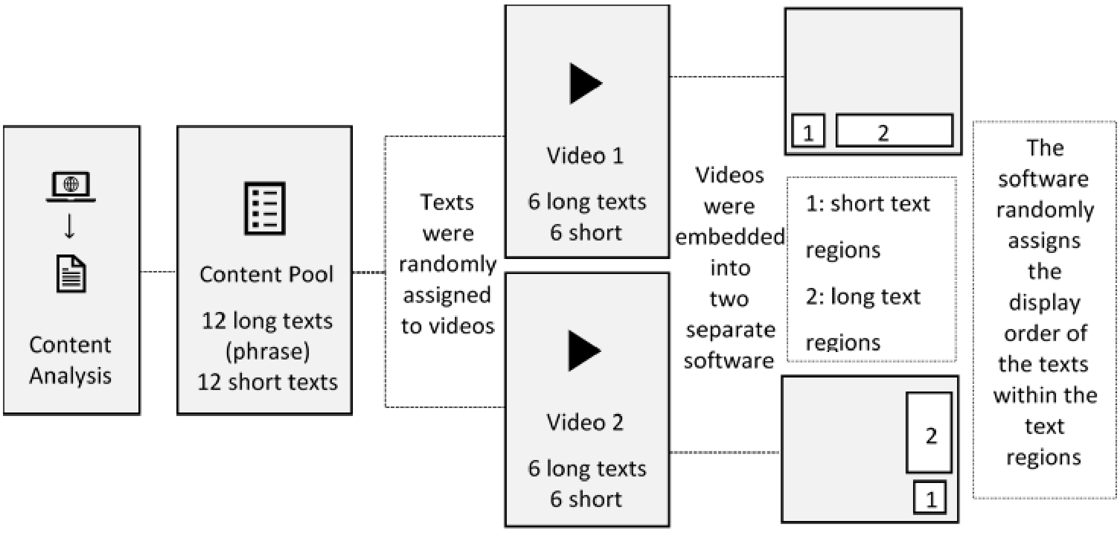

Figure 2. Design process of video materials

- Both of the videos have focused on the same topic of a financial conversation.
- Similar contents have been presented in the corresponding text regions. The contents on-screen texts were different from those of spoken ones.
- The same font and background colors have been selected for the informative texts.
- The size of "area of interest" for short-text regions is equal for each video, and so do for long-text regions.

To manipulate the contents inside the selected text regions, each video was embedded directly into a special software developed with Adobe Flash CS6 and programmed with Action Script 3.0. The following procedure was applied during the designing and programming process of the software (see **Figure 2**).

- First, the content being presented at the text regions was obtained from publicly available websites, then tabulated and summarized by the researcher.
- A content pool consisting of twelve long and twelve short texts were extracted from the summarized content repository. The number of characters in each corresponding text region was close to each other. In other words, while short texts included characters between 12 and 18, long texts included characters between 72 and 80.
- Both short and long texts in the pool were randomly assigned to corresponding text regions of the video materials. Consequently, each video had a pool of six short and six long texts to be presented in the related areas.
- When the letters (texts) on the screen replace with new ones, a disruption on the eye movement patterns occur (Rayner, 1998). To control the display change effect in the text regions, the texts were changed at ten-second intervals for each region in an orderly fashion. To do so, a sorting algorithm has been applied to determine the display order of the texts on the screen. This algorithm allows creating two new arrays with six indices for each text region of a video. Then, it randomly assigns six short texts to one array and six long texts to another array. When the software is started, the video appears in the full-screen mode and presents different short and long texts at each time. After ten seconds elapsed, the text at the following indices of the array is presented in the related region of the video. Text in each index of the arrays appears three times on the screen for each video.

The paper-based materials, on the other hand, included a prior knowledge test and a retention test. Before putting the materials into the final form, the questions were checked by an instructional technologist with PhD in terms of content and construct validity. Participants received these materials on individual 8.3 X 11.7-inch sheets of paper.

Prior knowledge (experience) test

This questionnaire consisted of two parts: (i) basic background information, and (ii) interest in financial issues and prior knowledge. The first part included several demographic questions, such as age, gender, educational background, and language. The second part aimed to identify whether participants interest in financial issues and to determine their prior knowledge in the economy. To do so, a 7-item knowledge checklist and a 5-item self-rating question (1= *very little*, 5= *very much*) were prepared. While preparing the questions, we inspired from the prior knowledge test developed by Moreno and Mayer (1999) to assess a different topic. The checklist included the following seven items: “I regularly follow economy news,” “I regularly read financial publications,” “I feel excited when I watch financial issues,” “I have purchased a share from the BIST (Borsa Istanbul) recently,” “I regularly follow the daily exchange rates,” “I know why The Turkish Central Bank has reduced Turkey’s growth forecast for the current year,” “I can explain waves in financial market.” Participants were asked to put a checkmark next to the items if the proposed statement was appropriate his or her. Additionally, on the self-rating question, participants were also asked to place a checkmark indicating their knowledge in finance (economy). The experience scores of participants were calculated by counting the number of domain-related items and adding the score addressing the level of knowledge in the economy from the self-evaluation question. Consequently, the total score that could be obtained from the questionnaire ranged from 1 to 12.

Retention test

There are two goals of multimedia learning, which are remembering and understanding (Mayer, 2001). Remembering refers to the ability to recognize or reproduce the presented material, and it can be measured through retention tests (Mayer, 2001). In the present study, a retention test was used. At the top of the sheet, the following instruction was attached: “Please do not select any option if you do not remember the answer of the given question.” The test contained eight multiple-choice and four true-false questions to measure to what extent participants remember the information presented in the text areas. Each multiple-choice question possessed a question stem and four options. One of the options provided under the stem for each question was related to the content presented with the texts on the videos. The test was prepared in a way that includes an equal number of questions from the information presented at the four text areas on the videos. Each correct answer on the retention test yielded one point; thus, the total score that can be obtained from the test was 12.

Data Analysis

IBM SPSS 21 was used for statistical analyses. To answer the first research question, the Pearson Correlation analysis was performed. Additionally, three sperate analyses were conducted to seek the second research question: comparison of (i) total fixation durations, (ii) fixation durations between short text regions, (iii) fixation durations between long text regions. Kolmogorov-Smirnov and Levene tests were used to check normality and homogeneity of variance assumptions before further analysis. When both of the assumptions were satisfied, one-way ANOVA was performed, if not so the Mann-Whitney U-test was conducted. The same procedure was also applied to the last research question.

RESULTS

The Relationship between Eye-fixation and Retention

The first research question concerns whether there is a relationship between the participants’ total fixation on the predetermined text regions and their retention scores received from the questions referring to these regions. Concerning this question, a correlation coefficient was computed between total fixation and retention scores. The results showed that there was a statistically positive correlation between the two variables, $r= .29$, $n= 61$, $p= .025$, which means that total eye-fixation duration shared 8% of the variability in total retention scores.

Table 1. Means and standard deviations of eye-fixation durations and retention scores

	Eye-fixation duration		Retention score	
Text positions	*M*	*SD*	*M*	*SD*
Bottom of the video	72.99	27.52	2.72	1.34
Right side of the video	100.72	21.46	2.77	1.16
Short text at the bottom of the video	15.79	9.70	.87	.72
Short text at the right side of the video	24.49	12.12	1.13	.67
Long text at the bottom of the video	57.20	21.73	1.85	1.03
Long text at the right side of the video	76.23	15.14	1.64	1.02

A further correlation coefficient was also computed between fixation duration and retention for four text regions separately. First, there was no significant relationship between the fixation duration and retention scores addressing to the short-text area presented at the bottom of the video, $r= .24$, $n= 61$, $p= .061$. Second, the eye-fixation duration of the long-text area presented at the bottom was not significantly correlated with the retention scores participants obtained from the questions related to this area, $r= .22$, $n= 61$, $p= .083$. On the other hand, a correlation for the data obtained from the short-text area presented at the right side of the video revealed a significant relationship between fixation duration and retention score, $r= .40$, $n= 61$, $p < .01$, which means 16% of the variance is explained. Finally, there was also a statistically significant correlation between two variables at the long-text area presented on the right side of the video, $r= .27$, $n= 61$, $p= .035$, which means there is only 7% of variance that is explained.

The Difference between Eye-fixation Duration at Two Videos

The second question seeks the mean difference between the total fixations of participants on the text positions presented at the videos. To test this research question, the Mann-Whitney U-test was conducted to compare the ranks for the total fixations in seconds on text positions presented at the bottom and on the right. The intersection of the first two rows and the first column of **Table 1** shows means and standard deviations for eye-fixation durations for two groups. The results indicated a statistically significant difference between the two layouts, U (122) = 2892, $Z = 5.28$, $p < .001$, $r= .48$, with the sum of rank 78.41 for the text on the right side and 44.59 for the text on the bottom. These results revealed that participants remained more fixed on the texts presented at the right side of the video than texts presented at the bottom of the video.

Additionally, a further one-way ANOVA was conducted to figure out the difference between fixation durations obtained from short text regions presented at the bottom of the video and the right side of the video. The analysis results revealed a statistically significant difference between two groups, F (1,121) = 19.20, MSE = 120.49, $p < .001$, $\eta^2 = .16$, which means that participants more fixed the short text regions presented at the right side of video than short text regions presented at the bottom of the video.

Furthermore, the Mann-Whitney U-test was also conducted to compare the ranks for the total fixations on long text positions presented at the bottom and right side of the video. The results indicated that the eye-fixation duration was significantly affected by the long-text position, U (122) = 2779.5, $Z = 4.71$, $p < .001$, $r= .43$, with the sum of rank 76.57 for the long text on the right side and 46.43 for the long text on the bottom. These results revealed that participants more fixed on the long texts presented at the right side of the video than long texts presented at the bottom of the video.

The Difference between Retention Scores at Two Videos

The third research question concerns whether the participants more remember the information presented at the bottom of the video than information presented at the right side of the video. The intersection of the first two rows and the second column of **Table 1** shows means and standard deviations for eye-fixation durations for two groups. The one-way ANOVA analysis revealed no significant difference between two layouts in participants' retention scores, F (1,121) =.05, MSE = 1.58, p = .83, $\eta^2 < .01$ which means that retention scores did not differ in terms of text layouts.

As a further test of remembering, retentions scores obtained from short text regions were compared between two layouts with one-way ANOVA. The results indicated a statistically significant difference between two layouts in retention scores, F (1,121) = 4.35, MSE = .48, p = .04, η^2 = .04. These results revealed that participants remembered more information from the short text regions presented at the right side of the video than short text regions presented at the bottom of the video.

A final one-way ANOVA was also performed to compare the retention scores participants obtained from the questions addressing the long text regions between two layouts. The analysis revealed no significant difference between two layouts in retention scores, F (1,121) = 1.32, MSE = 1.05, p = .25, η^2 = .01, which means that retention scores obtained from long text regions presented at the bottom of the video did not differ from text regions presented at the right side of the video.

DISCUSSION

The present study was designed to elaborate on the two text layouts frequently used at videos in terms of eye-fixation duration and retention. A one-shot case experimental design was employed to examine participants' total fixations on the predetermined text regions and retention scores they got from a retention test, including questions referring to the information presented in these regions. The results indicate a significant correlation between total eye-fixation duration and retention test scores. However, the magnitude of the correlation was different across four text regions and some (i.e., short-text area at the bottom and long-text area at the bottom) of which was not significant. The study also suggests that learners fixate more on the texts presented at the right sight of the screen than texts presented at the bottom. Additionally, contrary to the expectations, the study did not find any significant difference between the retention scores related to the text positions.

Previous studies have demonstrated a positive correlation between fixation and attention (Ferber & Murray, 2005; Henderson, 2007), and fixation and transfer (She & Chen, 2009), complying with the present results. However, a possible correlation between fixation duration and retention score was not analyzed before our study. Although we did not observe a strong relationship between these variables, a significant positive correlation has been observed between fixation and retention. This is an important result because meaningful learning requires a good retention ability (Mayer, 2001) and the goal of multimedia instruction is to help learners to transfer previously learned concepts to new situations. In this sense, the present finding draws our attention to the layout difference of informative texts on videos, and how viewers perceive these texts while interacting with videos. However, there are still many questions that need further investigation: e.g., which external factors are more related to the total fixation on instructional videos, and to what extent these factors predict the goals of the multimedia instruction.

Visual Processing of Texts on Videos

Early studies in eye movements have demonstrated that viewers direct their fixations towards informative and interesting regions of a scene (Henderson, 2003). The characteristics of the given task (e.g., reading, scene perception, visual search, etc.) also influence the gaze control behavior (Rayner, 2009). In a reading task, for example, fixation duration and location might be affected by several factors including text difficulty (e.g., being content or functional word, and number of the letters in the word), reading skills (e.g., beginner or dyslexic reader, and skilled reader) and the characteristics of the writing system (e.g., English or Chinese) (Rayner, 2009). In a scene perception, on the other hand, the salience of the material (Henderson, 2007; Schütz et al., 2011) and the cognitive knowledge structure of the viewer (Henderson, 2007) play a critical role in guiding eye movements. In this regard, the present study covers both of these task demands. In other words, it focuses on not only how viewers process text information in different forms (short and long texts), but also how they perceive these texts on a dynamic visualization (video). As a result, due to the complicated structure of the visual processing task and the lack of experimental control in one-shot case studies, several concerns may arise about the generalizability of findings. To lessen this concern, the following strategies were employed while designing the study: (a) selecting similar texts, (b) assigning the texts to the related regions of the video randomly, (c) designing the corresponding text regions in way that are visually close to each other, (d) using similar contents in each video, and (e) selecting low-experienced participants.

The content of each text in the corresponding text regions was similar to each other. Texts in short-text regions consisted of two lines: the names and the amount of parities, with the characters of 12 to 18. The count of characters in the long-texts, on the other hand, was in the range of 72 to 80. While long texts presented at the bottom of the video consisted of three lines, long texts presented at the right side of the video contained eight lines. The results of this study indicated that the total fixation on the texts presented on the right side of the video is statistically higher than that on the texts at the bottom and viewers more fixed on both short and long text regions presented at the right side of the video than corresponding text regions presented at the bottom of the video. A possible explanation of this is that learners may perceive any text presented at the bottom of a video as an additional information source.

On the other hand, texts presented on the right side of the screen may be associated with the main message delivered with the video. At this point, there might be a concern of the end-ofline effect in reading rather than layout differences. Concerning this issue, we attempted to control the length of the texts by selecting texts covering similar content classify, and containing characters close to each other. Nevertheless, the role of the line difference between long texts at the bottom of the video and long texts presented at the right side of the video might be arguable to some extent. However, any influence caused by the number of words per line might have revealed a divergent result regarding the total fixation since it has been shown that readers more fixate long words than short words (Reichle et al., 1998).

Kruger and Steyn (2014) have reported that the word count and the number of lines in a subtitle do not play a vital role in the processing of English subtitles in a lecture delivered in English. However, the length of a word (e.g., short or long) may affect the fixation durations. Readers may skip some words during a reading task depending on the number of words presented in the regions (Drieghe et al., 2005; Rayner et al., 2011). The probability of fixating is increased as the length of a word increased (Rayner, 2009). Similarly, the probability of fixating a word is decreased as the length of the word decreases (Rayner, 1998). Furthermore, the gaze control of readers is also affected by the predictability of the words in the text (Drieghe et al., 2005), and semantic relations between the words in the text regions and the main message of the video (Morris, 1994). The result of lower fixations on short text regions in the present study might be interpreted as the skipping effect in reading, as reported by previous studies. Therefore, it can be concluded that viewers more fixate on long text regions than short text regions in a scene perception task. However, a note of caution is due here because every individual has a unique way of looking at the components presented on a screen (Noton & Stark, 1971). In other words, viewers not only look at the salient regions or objects on the screen but also look at a part of the scene based on their cognitive knowledge structures (Henderson, 2007). However, the total fixation duration for the text regions might change based on the experience level of participants in the field (e.g., the stock market or economy).

Recognition of Informative Texts

One goal of multimedia representation is to present information in an efficient way (Mayer, 2001). The goal of multimedia instruction, on the other hand, is to deliver the information in a way that promotes both remembering and understanding (Mayer, 2001). Researches in the field have contributed to our understanding of how to design effective multimedia materials with the emphasis of extending the scope of these principles and establishing new ones (de Koning et al., 2018). At this point, the present study aims to improve our understanding of how to place informative texts on videos to enhance their memorability. This might be considered an important issue because instructional videos have become widespread in education (Bétrancourt & Benetos, 2018), and many students have already taken online courses supported by video materials. However, the present study did not find any significant difference between the total retention scores of the participants. Besides that, the study yielded diverse findings for information-retention for short and long texts. This was an unexpected result because the fixation duration of participants was higher for the text regions presented at the right side of the video than texts presented at the bottom. It is difficult to explain this result, but it may be argued that the fixation duration might not be a good predictor for learning outcomes.

Previous studies have suggested the use of spatial contiguity principle on static visualizations (Bétrancourt & Bisseret, 1998; Mayer, 1997; Mayer et al., 1995), and dynamic visualizations (Johnson & Mayer, 2012; Mayer & Anderson, 1992; Mayer & Moreno, 2002; Mayer & Sims, 1994). The presentation type of the multimedia content used in these studies meet on a common ground, demonstration of a fact including several procedures, such as a lesson about how car brakes work (Johnson & Mayer, 2012), operation of a bicycle (Mayer & Anderson, 1992), and how the human respiratory system work (Mayer & Sims, 1994). The role of the informative texts in these studies is to support the main message of the lesson by describing a concept or process illustrated on relevant visuals. On the other hand, the video materials used in the present study pursue neither a goal of teaching a “how-to” problem nor an instructional purpose. Instead, they merely cover the topic of recent financial issues and include informative texts addressing the news in the stock market and economy. Therefore, the results of the study might not be generalized to other domain-specific video materials.

Limitation of the Study and Implications for Further Studies

Due to several reasons, the results of the present study cannot be interpreted as conclusive, but suggestive. First, it might be relatively easier to manipulate the *salience* of the material. However, the *plans* and *values* may change depending on the characteristics of the viewer. In this sense, although the eye-tracking technique provides a good opportunity to gather saccade maps and fixation data (Catrysse et al., 2018), it may not give detail information about the intentions and the goals of learners. To understand where and why learners look a particular part of a scene; therefore, the eye-tracking technique needs to be accompanied by different strategies. In this study, we did not employ any additional method, so these restrictions should be considered while evaluating the results. Second, the content presented at the videos was about financial issues, was not an illustration of a fact requiring procedural learning demands. From this point of view, the results of this study may include cues about how to place informative texts on instructional videos delivered through online learning environments. However, the results might not be generalized for the presentation of all learning contents via video material, so further research is required to confirm and validate these results.

The present study has focused on the retention ability of the participants. However, the ultimate goal of multimedia instruction is to enable understanding, the transfer of the acquired information from the multimedia materials to new situations. In this regard, further studies are needed to understand whether informative texts, which are intelligible in isolation, presented at different positions of videos, play a critical role in the learning of different subjects.

An experienced user may perceive a scene different than those of novice users. Henderson (2007), for example, explains this situation by giving a simple example, learning the current time. If someone had prior experience with this task, most probably, he or she would look at the location where the time is provided. In this case, the salience of the materials in the visual field may not be effective in driving attention. Similarly, retention scores and fixation durations of experts in the field (e.g., stock market expert, etc.) may be different from unexperienced (novice) users. Therefore, further studies are required to understand how experienced users control their eye movements in such a task and whether there is a performance difference in retention scores between inexperienced users.

CONCLUSION

The study aimed to explore the role of the text positions placed at videos on eye-fixation duration and retention by applying the method of eye-tracking. A positive relationship has been observed between total fixation time and retention. However, the magnitude and the significance of the correlation is changed depending on the text positions. Additionally, viewers focus on texts longer if they are presented on the right side of the video than those positioned at the bottom of the video. Surprisingly, the study did not detect any difference between the retention scores learners obtained by giving correct answers to the questions referring to the texts presented at the bottom and on the right side of the instructional videos. Although the present study is based on a limited sample of content (financial issues), the findings suggest that while designing videos with on-screen texts, the right part of the screen is likely a better option if the prerequisite of the designer is to control the gaze behavior.

ACKNOWLEDGEMENT

Portions of this study were presented at the biennial conference of the 27th International Conference on Educational Science, April 2017, Antalya, Turkey. Additionally, the video materials used in the study were taken from https://www.bloomberght.com/video. The author also thanks three anonymous reviewers for their valuable and constructive comments on the manuscript.

REFERENCES

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202-232.

Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers and Education, 53*(4), 1339-1354. https://doi.org/10.1016/j.compedu.2009.06.017

Bétrancourt, M., & Benetos, K. (2018). Why and when does instructional video facilitate learning? A commentary to the special issue "developments and trends in learning with instructional video." *Computers in Human Behavior, 89*(August), 471-475. https://doi.org/10.1016/j. chb.2018.08.035

Bétrancourt, M., & Bisseret, A. (1998). Integrating textual and pictorial information via pop-up windows: An experimental study. *Behaviour & Information Technology, 17*(5), 263-273. https://doi.org/10.1080/014492998119337

Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior, 89*, 411-417. https://doi.org/10.1016/j. chb.2017.12.002

Boucheix, J. M., Gauthier, P., Fontaine, J. B., & Jaffeux, S. (2018). Mixed camera viewpoints improve learning medical hand procedure from video in nurse training? *Computers in Human Behavior, 89*, 418-429. https://doi.org/10.1016/j.chb.2018.01.017

Catrysse, L., Gijbels, D., & Donche, V. (2018). It is not only about the depth of processing: What if eye am not interested in the text? *Learning and Instruction, 58*, 284-294. https://doi.org/10.1016/j.learninstruc.2018.07.009

Chigona, W., Sonnet, H., Ritter, F., & Strothotte, T. (2003). *Shadows with a message BT - Smart Graphics* (A. Butz, A. Krüger, & P. Olivier, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg.

Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: an eye-tracking study. *Educational Technology Research and Development*, 1-19. https://doi.org/10.1007/s11423-018-9594-x

Cooper, D., & Higgins, S. (2015). The effectiveness of online instructional videos in the acquisition and demonstration of cognitive, affective and psychomotor rehabilitation skills. *British Journal of Educational Technology, 46*(4), 768-779. https://doi.org/10.1111/bjet.12166

de Koning, B. B., Hoogerheide, V., & Boucheix, J.-M. (2018). Developments and trends in learning with instructional video. *Computers in Human Behavior, 89*, 395-398. https://doi.org/10.1016/j.chb.2018.08.055

Deubel, H. (2008). The time course of presaccadic attention shifts. *Psychological Research, 72*(6), 630. https://doi.org/10.1007/s00426-008-0165-3

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 954-969. https://doi.org/10.1037/0096-1523.31.5.954

Duchowski, A. T. (2003). *Eye tracking methodology: Theory and practice*. London: Springer-Verlag.

Ferber, S., & Murray, L. J. (2005). Are perceptual judgments dissociated from motor processes? - A prism adaptation study. *Cognitive Brain Research, 23*(2-3), 453-456. https://doi.org/10.1016/j.cogbrainres.2004.11.007

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior, 89*, 465-470. https://doi.org/10.1016/j.chb.2018.07.015

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review, 16*, 41-67. https://doi.org/10.1016/j.edurev.2015.06.001

Gersch, T. M., Kowler, E., Schnitzer, B. S., & Dosher, B. A. (2009). Attention during sequences of saccades along marked and memorized paths. *Vision Research, 49*(10), 1256-1266. https://doi.org/10.1016/j.visres.2007.10.030

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511-525. https://doi.org/10.1016/j.learninstruc.2006.10.001

González-Gómez, D., Jeong, J. S., Airado Rodríguez, D., & Cañada-Cañada, F. (2016). Performance and perception in the flipped learning model: An initial approach to evaluate the effectiveness of a new teaching methodology in a general science classroom. *Journal of Science Education and Technology, 25*(3), 450-459. https://doi.org/10.1007/s10956-016-9605-9

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*, 188-193. https://doi.org/10.1016/j.tics.2005.02.009

Hayhoe, M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1), 49-63. https://doi.org/10.1167/3.1.6

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498-504. https://doi.org/10.1016/j.tics.2003.09.006

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*(4), 219-222. https://doi.org/10.1111/j.1467-8721.2007.00507.x

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences, 21*(1), 15-23. https://doi.org/10.1016/j.tics.2016.11.003

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 210-228. https://doi.org/10.1037/0096-1523.25.1.210

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57*(6), 787-795. https://doi.org/10.3758/BF03206794

Hoogerheide, V., van Wermeskerken, M., van Nassau, H., & van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Computers in Human Behavior, 89*, 457-464. https://doi.org/10.1016/j.chb.2017.11.012

Howard, E., Meehan, M., & Parnell, A. (2018). Live lectures or online videos: Students' resource choices in a first-year university mathematics module. *International Journal of Mathematical Education in Science and Technology, 49*(4), 530-553. https://doi.org/10.10 80/0020739X.2017.1387943

Huh, D., Kim, J. H., & Jo, I. H. (2019). A novel method to monitoring changes in cognitive load in video-based learning. *Journal of Computer Assisted Learning, 35*(6), 721-730. https://doi.org/10.1111/jcal.12378

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology. Applied, 18*(2), 178-191. https://doi.org/10.1037/a0026923

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105-120. https://doi.org/10.1002/rrq.59

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328. https://doi.org/10.1068/p2935

Leo, J., & Puzio, K. (2016). Flipped instruction in a high school science classroom. *Journal of Science Education and Technology, 25*(5), 775-781. https://doi.org/10.1007/s10956-016-9634-4

Love, B., Hodge, A., Grandgenett, N., & Swift, A. W. (2014). Student learning and perceptions in a flipped linear algebra course. *International Journal of Mathematical Education in Science and Technology, 45*(3), 317-324. https://doi.org/10.1080/0020739X.2013.822582

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. (2001). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167-171. https://doi.org/10.1016/j. learninstruc.2009.02.012

Mayer, R. E., & Anderson, R. B. (1992). The Instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*(4), 444-452. Retrieved from https://www.learntechlib.org/p/147036

Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14*(1), 87-99. https://doi.org/10.1023/A:1013184611077

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

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory*. https://doi.org/10.1017/CBO9780511844744.009

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words. Extensions of a dual coding theory of multimedia. *Journal of Educational Psychology, 86*(3), 389-401.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43*(1), 31-41. https://doi.org/10.1007/BF02300480

Merkt, M., Ballmann, A., Felfeli, J., & Schwan, S. (2018). Pauses in educational videos: Testing the transience explanation against the structuring explanation. *Computers in Human Behavior, 89*, 399-410. https://doi.org/10.1016/j.chb.2018.01.013

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368. https://doi.org/10.1037/0022-0663.91.2.358

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 92-103. https://doi.org/10.1037/0278-7393.20.1.92

Nadelson, L. S., Scaggs, J., Sheffield, C., & McDougal, O. M. (2015). Integration of video-based demonstrations to prepare students for the organic chemistry laboratory. *Journal of Science Education and Technology, 24*(4), 476-483. https://doi.org/10.1007/s10956-014-9535-3

Nagy, J. T. (2018). Evaluation of online video usage and learning satisfaction: An extension of the technology acceptance model. *International Review of Research in Open & Distance Learning, 19*(1), 160-184.

Noton, D., & Stark, L. (1971). Eye movements and visual perception. *Scientific American, 224*(6), 34–43. Retrieved from http://www.jstor.org/stable/24922750

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers and Education, 53*(2), 445-453. https://doi.org/10.1016/j.compedu.2009.03.002

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford, UK: Oxford University Press.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400*(6741), 233-238. https://doi.org/10.1038/22268

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. Journal of Experimental Psychology. *Human Perception and Performance, 37*(2), 514-528. https://doi.org/10.1037/a0020990

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157. https://doi.org/10.1037/0033-295X.105.1.125

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24*(2), 226-237. https://doi.org/10.1002/acp.1554

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision, 11*(5), 1-30. https://doi.org/10.1167/11.5.9

She, H.-C., & Chen, Y.-Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers & Education, 53*(4), 1297-1307. https://doi.org/10.1016/j.compedu.2009.06.012

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York, NY: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261-292. https://doi.org/10.1007/s10648-019-09465-5

Tang, F., Chen, C., Zhu, Y., Zuo, C., Zhong, Y., Wang, N., Zhou, L., Zou, Y., & Liang, D. (2017). Comparison between flipped classroom and lecture-based classroom in ophthalmology clerkship. *Medical Education Online, 22*(1), 1395679. https://doi.org/10.1080/10872981. 2017.1395679

Traphagan, T., Kucsera, J. V, & Kishi, K. (2010). Impact of class lecture webcasting on attendance and learning. *Educational Technology Research and Development, 58*(1), 19-37. https://doi.org/10.1007/s11423-009-9128-7

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247-262. https://doi.org/https://doi.org/10.1006/ijhc.2002.1017

van der Meij, H. (2017). Reviews in instructional video. *Computers and Education, 114*, 164-174. https://doi.org/10.1016/j.compedu.2017.07.002

van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication, 60*, 205-228. Retrieved from https://www.ingentaconnect.com/content/stc/tc/2013/00000060/00000003/art00004

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95-99. https://doi.org/10.1016/j.learninstruc.2009.02.009

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89*, 430-438. https://doi.org/10.1016/j.chb.2017.11.038

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management, 43*(1), 15-27. https://doi.org/10.1016/j.im.2005.01.004

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human Computer Studies, 45*(1), 59-74. https://doi.org/10.1006/ijhc.1996.0042

Correspondence: Hamza Polat, Department of Computer Education and Instructional Technology, Erzincan Binali Yıldırım University, Erzincan 24100, Turkey. E-mail: hpolat@erzincan.edu.tr