International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Timeline-Anchored Comments in Video-Based Learning: The Impact of Visual Layout and Content Depth

Yue Chen, Qin Gao & Ge Gao

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2021.1976505

Timeline-Anchored Comments in Video-Based Learning: The Impact of Visual Layout and Content Depth

Yue Chen[a], Qin Gao[a], and Ge Gao[b]

[a]Department of Industrial Engineering, Tsinghua University, Beijing, China; [b]The College of Information Studies, the University of Maryland, College Park

ABSTRACT

Many video sites or learning platforms allow real-time chatting or asynchronous commenting on specific time points during video lectures. Comments, as user-generated knowledge, facilitate social interaction but also affect cognitive learning. The visual layout of these comments can affect learners' attention and learning, but the effect has rarely been studied. This study compares two common layouts (embedded vs. separated) and considers the content depth of comments through a laboratory eye-tracking experiment involving 40 participants. The results suggest that, with both layouts, learners switched attention to the comments every 10 seconds and stayed focused for 1.3 seconds on average before returning attention to the video. With an embedded layout, learners switched attention more frequently to the comments and remembered more surface-level comments. With a separate layout presenting deep-level comments, learners searched for information faster and performed better on open-book quizzes. We outline the design implications of using timeline-anchored comments to promote online learning.

1. Introduction

Video-based learning has become a common way for online learners to obtain knowledge both formally (e.g., from lectures provided by schools or massive open online courses) and informally (e.g., from videos of interests shared on YouTube). Particularly during the recent pandemic, online learning via prerecorded videos (e.g., most massive online open course platforms and distant learning platforms), online conferencing systems (e.g., Zoom), or live streaming platforms (e.g., general-purposed live streaming sites such as Facebook Live or educational live streaming sites) have been adopted by schools worldwide (Li & Lalani, 2020). Compared with offline classroom learning, a major limitation of video-based learning is the lack of social interaction, which is beneficial to meaningful learning (Akyol & Garrison, 2011; Baker, 2010; Garrison & Arbaugh, 2007; Jung & Lee, 2018; Ke & Kwak, 2013; Rovai, 2002).

A major approach to promote social interaction in video-based learning is to integrate and synchronize comments or discussions specific to certain time points of a video. In online conferencing systems and live streaming platforms, such synchronized discussion is achieved by incorporating live chatting functions. For recorded videos, including the recorded live-streamed videos for replay, such synchronization can be achieved by anchoring viewers' comments to the playback time on the video when the comments are posted. These comments will be seen by viewers who play the video later. Such integration and synchronization of the video content and time-point-specific comments creates a real-time (as in live streaming and online conferencing) or pseudo real-time (as in watching prerecorded videos) interaction experience among viewers (Johnson, 2013). Such timeline-anchored comments can be displayed either in a separate manner, i.e., comments are displayed in a sidebar next to the video, or embedded into the video as scrolling text or flying subtitles, as shown in Figure 1.

Several studies have explored the impact of timeline-anchored commenting on the learning experience and performance and found that the functions allow learners to express their immediate opinions or feelings, promote social interactions, and create a lively atmosphere (Chen et al., 2019; Lee et al., 2015, 2015; Yao et al., 2017; Yardi, 2006). In addition to socioemotional messages, viewers of lecture videos often share content-related messages, such as asking questions about the content, answering others' questions, and sharing content-related information or opinions (Chávez et al., 2016; Chen et al., 2019; Lin et al., 2018). Such content-related comments can serve as an additional source of information and knowledge for learning. The enriched data, on the other hand, may put higher demand on learners' attention distribution and management and result in a greater workload (e.g., Chen et al., 2017, 2019; Yardi, 2006).

How learners distribute their attention between the video and the discussions is strongly affected by the visual layout design of the time-point-specific comments. With the embedded layout, discussions are more visually salient and proximate to related video content (Pashler, 1988; Wickens & McCarley, 2007) so that learners can integrate information from comments and videos with less effort (Johnson & Mayer, 2012; Mason et al., 2013; O'Keefe et al., 2014). Some

CONTACT Qin Gao gaoqin@tsinghua.edu.cn;gaotsin@gmail.com Department of Industrial Engineering, Tsinghua University, Beijing, China

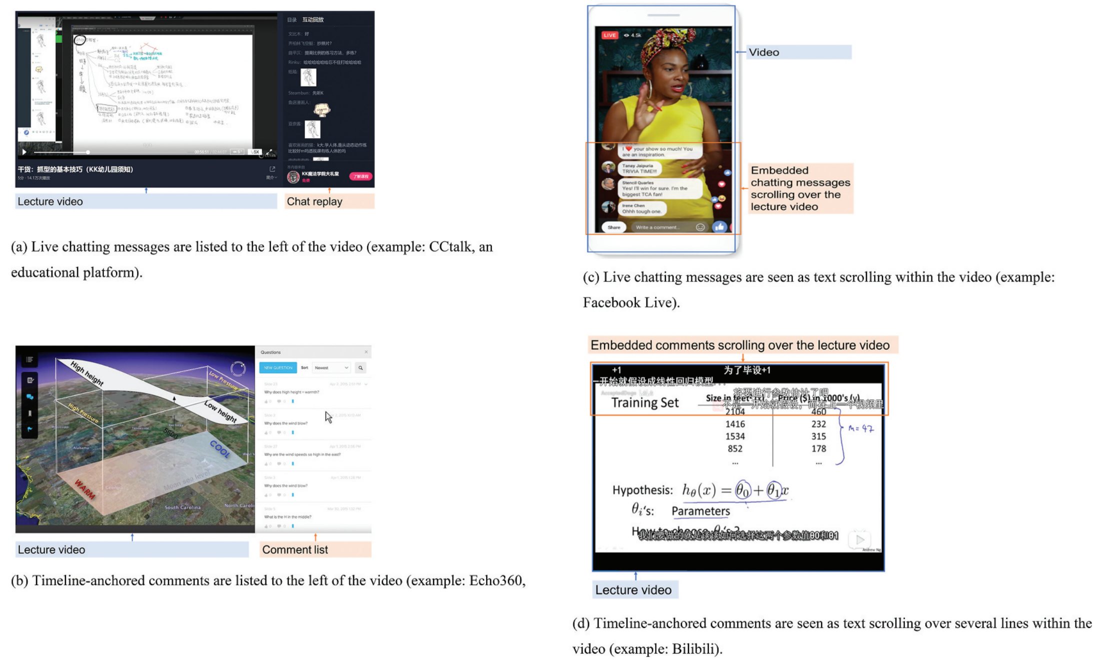


Figure 1. Screenshots of platforms that support live chatting or timeline-anchored commenting. (a) Live chatting messages are listed to the left of the video (example: CCtalk, an educational platform). (b) Timeline-anchored comments are listed to the left of the video (example: Echo360, an educational platform). (c) Live chatting messages are seen as text scrolling within the video (example: Facebook Live). (d) Timeline-anchored comments are seen as text scrolling over several lines within the video (example: Bilibili).

researchers have implied that such comments can work as in-video prompts that prevent learners' thoughts from wandering (Makransky et al., 2019; Shin et al., 2018). However, these comments are also scattered, transient, and hard for learners to catch up within the short time they are shown on screen (Leahy & Sweller, 2011; Singh et al., 2012; Sweller, 2011). In contrast, the separate layout usually lists all the comments next to the video, which can be scrolled and searched at any time. The list may work as structural cues or keywords in search and navigation tasks (Chen et al., 2019; Hsu & Schwen, 2003; Lu et al., 2018; Ma & Cao, 2017; Yousef et al., 2015). No empirical studies have been conducted to investigate the influence of different layout designs on learners' attention allocation and cognitive workload.

To help designers make informed decisions between the two design options, the present study aimed to investigate how different visual layouts of timeline-anchored commenting affect learners' attention allocation, cognitive workload, and learning experience. We compared the two main layouts for timeline-anchored comments (embedded or separate) with the consideration of the depth of information, i.e., how much cognitive effort is required to process the content of comments (deep or surface). We conducted a mixed design experiment involving 40 participants. They were asked to watch lecture videos with timeline-anchored comments and take quizzes in the lab. Eye tracking was adopted because it provides objective, continuous, and straightforward measures of how people direct their attention. In addition, participants' mind wandering and mental workload were measured through post hoc questionnaires.

2. Literature review

2.1. Attention allocation between video and comments

Video-based learning with comments requires learners to switch attention or eye gazes between two information sources: the video and the comments. The visual layout of the discussion can affect (1) how frequently learners switch their attention between the two sources and (2) how much attention they allocate to each source. Eye tracking has typically been adopted to examine such visual attention allocation. Compared with self-reporting and behavioral measures, eye tracking provides more objective, continuous, and straightforward measures of how people direct their attention. The frequency of gaze transitions between multiple informational sources was used as an indicator of learners' cognitive integration of information (Krejtz et al., 2016; Mason et al., 2015; O'Keefe et al., 2014), and the number of gaze fixations in a particular area reflects the amount of attention allocated to that area (Scheiter & Eitel, 2017; Schmidt-Weigand et al., 2010).

According to the salience, effort, expectancy, and value (SEEV) model (Wickens & McCarley, 2007), the probability that attention is attracted by or that the eyes move to an area depends on (a) the visual salience of the area; (b) the effort needed to access the information (i.e., the physical distance); (c) the user's expectancy of finding relevant information in the area; and (d) the perceived value of the information. Theoretically, the optimal attention allocation should be determined by the product of expectancy and value, which defines the expected value of looking at an object. In reality, however, people's attention allocation is modified by the other two nuisance factors. Empirical research of learning has also suggested that learning materials with higher visual salience (such as brighter colors and scrolling texts) attract more frequent gazing and visual attention (Jamet, 2014; Scheiter & Eitel, 2015).

2.2. Mind wandering

Attention can drift away from external tasks and perceptual input toward a more private, internal stream of consciousness (Mcmillan et al., 2013). Such mind wandering during learning can impair comprehension (Schooler et al., 2004; Smallwood et al., 2008). It occurs as frequently in online video learning as it does in classroom learning (Lindquist & McLean, 2011; Risko et al., 2012). According to the control failure hypothesis (McVay & Kane, 2010), mind wandering occurs when task-unrelated thoughts defeat task-related thoughts in competition for limited cognitive resources (Smallwood, 2010). One approach to help learners keep task-related thoughts in mind is incorporating simple activities that occur either continuously or randomly, such as interpolated testing, note taking (Szpunar, Khan et al., 2013), and prompting in video lectures (Shin et al., 2018). As with prompting in videos, timeline-anchored content-related comments can work as a task-relevant reminder to reduce mind wandering (Shin et al., 2018; Szpunar, Khan et al., 2013).

2.3. Cognitive learning outcomes

The effects of layouts on attention further influence learning performance and workload. This study investigates learners' performance on two common learning tasks: 1) knowledge retention and 2) knowledge search and integration. Knowledge retention is basic and fundamental to any kind of cognitive learning (Bloom, 1956; Krathwohl, 2002), requiring learners to store relevant information in their long-term memory and recall it later. To depict the learning process requiring the integration of multiple sources of information, a frequently used method is the spatial and temporal contiguity effect in Mayer's cognitive theory of multimedia learning (Ginns, 2006; Mayer, 2009), in which learners have to split their attention, which demands a considerable amount of working memory (Ayres & Sweller, 2005; Sorden, 2005). To reduce this split-attention effect, the interface needs to display information in high-contiguity or more proximate ways to integrate all relevant information (Wickens & McCarley, 2007).

Knowledge search and integration require learners to search, navigate, review, and integrate knowledge to construct meaning. Searching and revisiting knowledge that has been learned before is common in real-life scenarios (Eilertsen & Valdermo, 2000). Searching is necessary when learners need to comprehend and integrate knowledge from multiple sources and review the knowledge later. Such tasks have been found to be facilitated by visual cues that provide a map or overview for all of the learning materials (Hsu & Schwen, 2003; Li et al., 2013; Lu et al., 2018).

2.4. Content depth of comments

Learners' retention of comments and relevant knowledge can be affected by the extent of how much effort the comments require to process. The required effort can be roughly represented by the content depth (Hara et al., 2000; Henri, 1992) of the comments. Surface-level comments are defined as the simple restatement of a single piece of information from the lecture, for example, repeating some phrases as notes or the translation of terms (Hara et al., 2000; Henri, 1992), whereas deep-level comments are those that cannot be directly obtained from the lecture but require authors to add their own thoughts that are not expressed in the instructional material. The added content may include linking various facts, ideas, and experiences, making inferences from the evidence, comparing different knowledge points, and summarizing salient points.

Deep-level comments may trigger learners' deep cognitive processing, which means they engage in meaningful learning, organize their ideas, compare knowledge with existing knowledge in a wider context, and use life experience to understand concepts (Catrysse et al., 2018; Dolmans et al., 2016; Offir et al., 2008; Vermunt & Vermetten, 2004). These deep cognitive processing behaviors may all contribute to a better comprehension of the knowledge (Dinsmore & Alexander, 2016; Sins et al., 2008; Vansteenkiste et al., 2004). However, deep-level comments generally require more time to process; they are often longer and more integrated (Chávez et al., 2016; Hara et al., 2000; Henri, 1992) and trigger deep cognitive processing that is more time-consuming than surface processing (Lockhart & Craik, 1990).

3. Hypotheses

This study investigates the effects of two variables of timeline-anchored commenting on learners' attention allocation, mind wandering, and cognitive learning outcomes. One is the visual layout (embedded or separate), and the other is the content depth (deep or surface). Figure 2 summarizes the relationships of these variables and the hypotheses formed in this section.

As suggested by the SEEV model and previous empirical eye-tracking studies, the frequency of attention transitions increases when the distance between multiple information sources is closer (Bauhoff et al., 2012; Johnson & Mayer, 2012). The embedded layout displays comments as text scrolling over the video, which can achieve a higher visual salience (Abrams & Christ, 2003; Pashler, 1988) than the separate layout. It also presents the video and the comments more proximately and reduces the effort needed to access the comments. Therefore, we propose the following hypotheses:

Hypothesis 1. ***Attention transitions***: When learning with the embedded layout, learners switch attention more frequently between the video and the discussion.

Hypothesis 2. ***Gaze fixations***: When learning with the embedded layout, learners have more eye gaze fixations on the discussion.

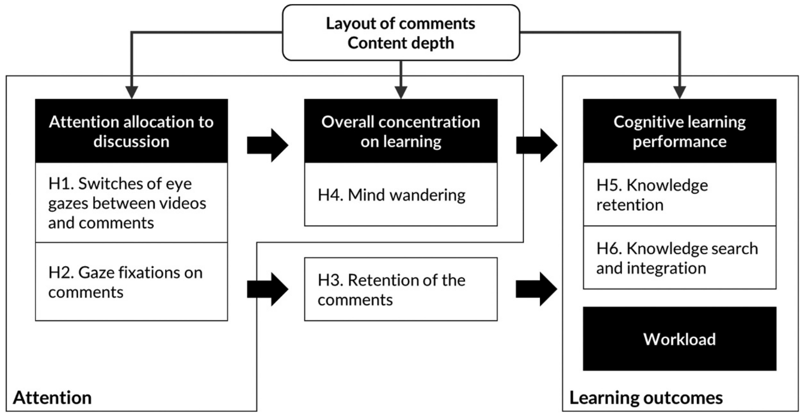


Figure 2. Relationships of independent variables, dependent variables, and hypotheses.

More gazes or attention to comments can be associated with greater retention of comments (Chun & Turk-Browne, 2007), especially task-oriented comments. However, comments in the embedded layout are dynamic and transient and only available for several seconds on the screen, making them difficult to assimilate (Leahy & Sweller, 2011; Singh et al., 2012; Sweller, 2011). Moving text is also more difficult to read than static text (Uetsuki et al., 2017). Learners may have insufficient time to process deeper transient information or miss it entirely because the input of moving elements is overwhelming (De Koning et al., 2010). Therefore, deep-level comments in the embedded layout may be harder for learners to catch up. Therefore, we propose the following hypotheses:

Hypothesis 3a. ***Retention of comments – layout***: When learning with the embedded layout, learners memorize more task-oriented comments.

Hypothesis 3b. ***Retention of comments – interaction effect***: The effect of layout on the retention of comments is greater when the comments are at the surface level than at the deep level.

As suggested in the Literature Review, comments embedded in the video may work as in-video prompts, which have been found to reduce mind wandering (Shin et al., 2018; Szpunar, Khan et al., 2013). Embedded in-video comments are especially effective because they are arguably more noticeable. Previous eye-tracking evidence also suggests that the embedded layout in higher spatial contiguity may reduce the time spent on irrelevant information (Leng et al., 2016; Makransky et al., 2019). Therefore, we propose the following hypothesis:

Hypothesis 4. ***Mind wandering***: When learning with a separate layout, learners' minds wander more.

As Hypothesis 1 suggests, the embedded layout may increase attention transitions between videos and comments, which may further promote knowledge retention (Johnson & Mayer, 2012; Mason et al., 2013; O'Keefe et al., 2014). Hypothesis 2 suggests that embedded comments are presented in a timely manner with a higher visual salience that attracts learners' attention more (Wickens & McCarley, 2007). This may invite passive participation in discussions. Empirical studies have found that passive participation in forums predicts learning performance even more strongly than active participation, such as posting threads (Brooker et al., 2018; Chiu & Hew, 2018; Wise & Cui, 2018). Therefore, we propose that the embedded layout probably helps learners recall relevant knowledge.

Hypothesis 5. ***Knowledge retention***: When learning with the embedded layout, learners recall more knowledge.

Timeline-anchored comments in a separate layout can serve as labels or keywords for an overview and thus may improve the performance of knowledge searching and comprehension (Hsu & Schwen, 2003; Li et al., 2013; Lu et al., 2018). Learning materials with deep-level information can help learners better comprehend the knowledge at a deeper level (Dinsmore & Alexander, 2016; Sins et al., 2008; Vansteenkiste et al., 2004); thus, deep-level comments may be more helpful for comprehension. The layout and depth of comments may also have an interactive effect on knowledge search and integration based on Hypothesis 3b. Therefore, we propose the following hypotheses:

Hypothesis 6a. ***Knowledge search and integration – layout:*** When learning with a separate layout, learners perform better in knowledge search and integration.

Hypothesis 6b. ***Knowledge search and integration – depth***: When learning with deep-level comments, learners perform better in knowledge search and integration.

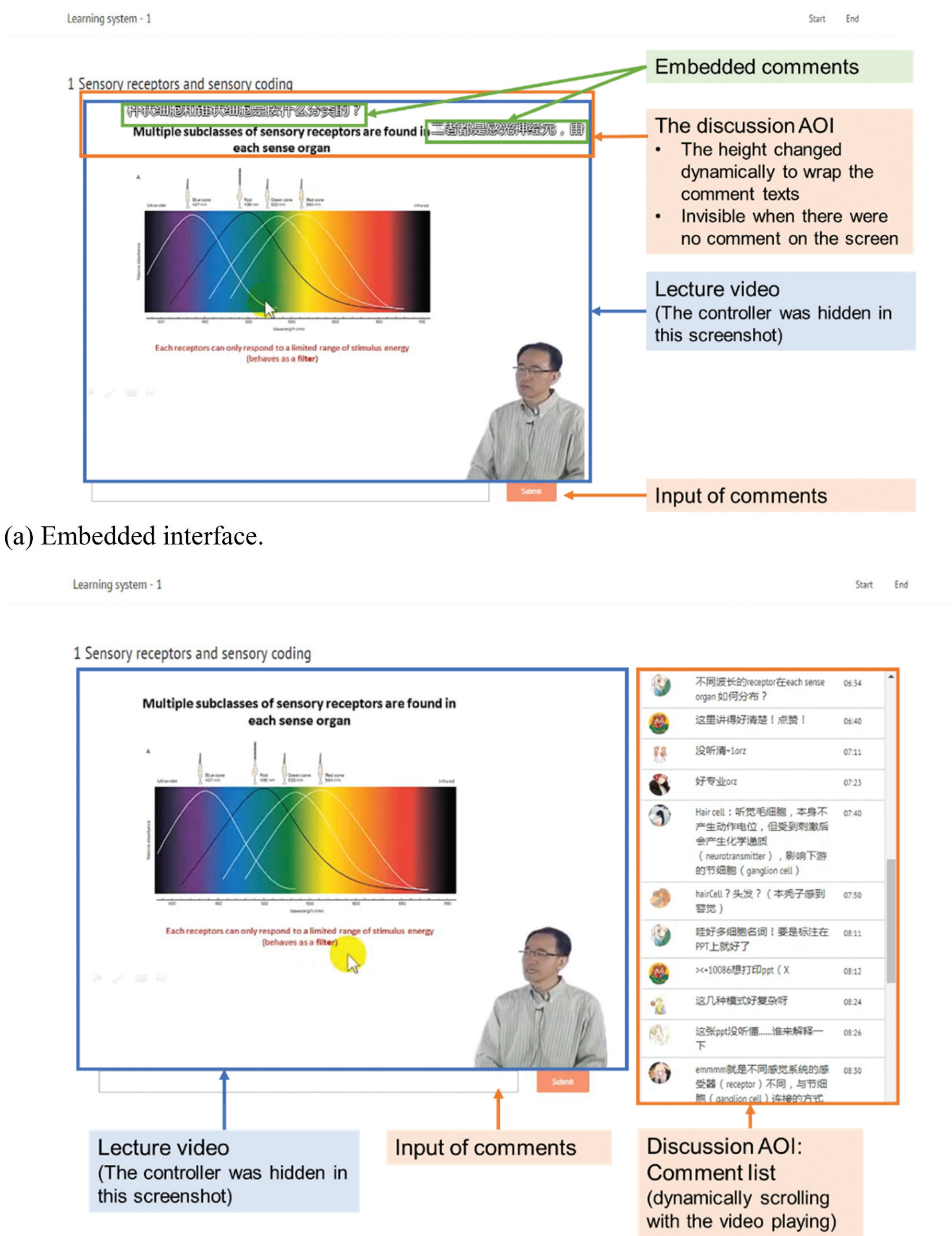
Hypothesis 6c. ***Knowledge search and integration – interaction effect***: The effect of layout on the performance of knowledge search and integration is stronger when the comments are at a surface level rather than at a deep level.

4. Method

4.1. Experiment design

We conducted an experiment using a mixed-group design. The independent variables were layout and depth of comments. Layout (embedded/separate) was a between-group variable, whereas depth (deep/surface) was a within-group variable. Each participant watched two lecture videos that contained timeline-anchored comments.

In the embedded condition, comments were presented by text scrolling over the video as it played (Figure 3a). Comments appeared in the upper quarter of the video area only. In the separate condition, comments were listed in a right sidebar (Figure 3b). The comment list automatically scrolled as the video played, with the current comment scrolling to the top of the sidebar. If the participant placed the cursor on the sidebar, the list would stop scrolling, so the



(b) Separate interface.

Figure 3. Screenshots of the platform. (a) Embedded interface. (b) Separate interface.

comments could be read carefully. If the participant clicked on a comment, the video would skip to the time point to which this comment was anchored. Each comment in the sidebar list indicated the author's avatar, the content of the comment, and the video time point to which the comment was anchored.

4.2. Learning materials: Lecture videos and comments

The lecture videos were chosen from a neuroscience course titled *Perception and Action: System Neuroscience*,[1] which was available on XuetangX.com. Because we had a two-level within-group variable, two videos were selected. Video 1 concerned sensory receptor and sensory coding, and video 2 concerned eye movements and extraocular muscles. Both videos lasted for nine to 10 minutes.

Three researchers evaluated the videos and declared them suitable for the experiment for the following reasons. First, the course was declarative and involved many objective facts and knowledge to remember and integrate. Therefore, it was easier to assess participants' attention, information seeking and integration, remembering, and understanding. Second, the content of the two videos was not in chronological order and was relatively independent. Both videos conveyed lectures in the same style by the same instructor and lasted for similar time periods. Therefore, it was convenient for counterbalancing conditions. Third, participants were not familiar with the content of the videos, but they could understand it (see the section on the participants).

Comments were collected from a previous study (Chen et al., 2019), in which four researchers and 28 participants posted timeline-anchored comments and threads. In this study, we decomposed those threads into multiple comments. We gathered 67 comments (of 1,622 characters) in the first video and 63 comments (of 1,506 characters) in the second. Comments consisted mainly of Chinese characters, with some biological terminology in English.

We invited six graduate students to independently categorize the comments into three groups: deep-level, surface-level, and social-oriented (Hara et al., 2000; Henri, 1992). Socially oriented timeline-anchored comments require little deep information processing by either authors or readers, and they have been found to have little effect on cognitive learning (Lee et al., 2015). Comments with task-oriented content that focus on the learning materials comprise the majority of content in online learning discussions (Chávez et al., 2016) and are considered useful by most learners. Task-oriented needs are the primary reason for learning from a lecture with timeline-anchored comments (Lin et al., 2018). Therefore, we concentrated on comparing deep – and surface-level task-oriented comments. The Fleiss' kappa coefficients of the categorizations in both videos were 0.66. Inconsistent categories of comments were discussed. Finally, we gathered 14 deep-level comments (659 characters), 35 surface-level comments (733 characters), and 18 social-oriented comments (230 characters) in video 1 and 20 deep-level comments (755 characters), 25 surface-level comments (509 characters), and 18 social-oriented comments (242 characters) in video 2 (see Table 1 for details).

Both deep-level and social-oriented comments were displayed in the deep-level condition, and both surface-level and social-oriented comments were displayed in the surface-level condition. We display the social-oriented comments in all the conditions because it was not the case that all the comments were task-oriented. Social-oriented comments may increase social interactivity without negatively impacting learning outcomes (Lee et al., 2015). In all the conditions and in both videos, the characters of social-oriented comments comprised approximately 30% of the total number of characters in the comments.

It is worth noting that the workload distraction level was affected by the number of comments. In this experiment, comments were set to contain approximately 100 characters per minute (mainly in Chinese). This was not distracting and

Table 1. Comments during the experiment.

Type of comments	Numbers and examples			
Task-oriented deep-level comments	Video 1 Number of comments	Number of characters	Video 2 Number of comments	Number of characters
	14	659	20	755
	Examples: "We can control the orientation and the amplitude of a saccade, but we cannot control the speed, which depends on the distance between the start and the end of the saccade." (Summarize information from different slides.) "This function is similar to the focusing system of optical cameras." (Add the author's thoughts.)			
Task-oriented surface-level comments	Video 1 Number of comments	Number of characters	Video 2 Number of comments	Number of characters
	35	733	25	509
	Examples: "May I request the translation of terms in this slide?" (Ask for simple explanation.) "I didn't know that the sense of pain is so complex and that it involves three types of receptors." (Repeat the slide content.)			
Social-oriented comments	Video 1 Number of comments	Number of characters	Video 2 Number of comments	Number of characters
	18	230	18	242
	Examples: "This part is introduced clearly. Thanks!" "It's hard for me to keep up with the professor because he is talking in a mixed language of English and Chinese."			

was acceptable in the learning context. In real-world scenarios, viewers may be overwhelmed by a huge number of comments, but any platforms with timeline-anchored commenting (such as Bilibili and Niconico) provide functions for users to control the number of visible comments on the screen.

4.3. Platform

The experimental platform was web-based and was developed by the Django 1.10.4 framework. It recorded clicking behaviors on the video in a database with timestamps of lecture videos. In the experiment, the platform was shown in the Chrome browser and presented the screen of the SMI iViewX RED60 eye tracker.

4.4. Measures

4.4.1. Attention transitions

Attention switch patterns were measured by the SMI iViewX RED60 eye tracker and analyzed using BeGaze software. To study the learners' switches between video and comments, we extracted features of the area of interest (AOI, i.e., the area of comments in this study). We manually defined AOIs in BeGaze. In the embedded condition, the AOI was a rectangle as wide as the lecture video. When the video was playing, the height of the AOI changed dynamically to accommodate the comments. The AOI was neither visible nor counted if there was no comment on the screen. In the separate condition, the AOI was simply the comment list. Attention switch patterns can be indicated by the following two AOI features. **Glance frequency** is the count per second each time a fixation hits the AOI if it is not hit before. **The average glance duration** is the mean value of the duration of each glance, calculated by the saccade duration of entering the AOI plus the length of time before the eyes begin to leave.

4.4.2. Gaze fixations and retention of comments

The fixation count on comments measured how comments attracted participants' visual attention (Scheiter & Eitel, 2017; Schmidt-Weigand et al., 2010), that is, the number of fixations on the AOI. Participants' attention to task-oriented comments was measured by the **performance of a memory test question**: "Did the following comments appear in this video?" Under each condition, the question contained five or six options for participants to judge. In each trial, the participants encountered one memory test question for the task-oriented comments. The options changed according to the condition and video. Performance was measured by the number of comments that were correctly judged by the participant.

4.4.3. Mind wandering

Mind wandering was measured by a four-item, 5-point Likert scale adapted from Cegala (1981). Cronbach's alpha was 0.72. Many studies have measured mind wandering by self-report probe methods (Weinstein, 2017). However, such methods would have proven highly distracting in the short-term learning session in this experiment. Therefore, participants reported mind wandering after the learning session was completed. Mind wandering was also found to be positively associated with the AOI feature average fixation duration, which was the mean value of each fixation duration (Reichle et al., 2010). Two global features of eye tracking also reflect mind wandering. Saccade frequency is the count of saccades (rapid eye movements between fixations) per second. The average saccade velocity is the mean value of the saccade angular velocity. Saccades were found to be less frequent and slower during mind wandering (Uzzaman & Joordens, 2011).

4.4.4. Learning performance and workload

Learning performance was measured by quizzes after the learners viewed the videos. For each video, we designed both closed-book and open-book quizzes. To assess knowledge recall, closed-book quizzes consisted of questions that required learners to recognize and recall the knowledge in the lecture videos. The closed-book quiz for video 1 had three multiple-choice questions, five true-or-false questions, and a matching question containing nine types of sensory and five receptors. The closed-book quiz for video 2 had three multiple-choice questions, eight true-or-false questions, and three fill-in-the-blank questions. The **score for the closed-book quiz** was calculated by the proportion of correct answers relative to the total number of questions.

To assess knowledge search and integration, an open-book quiz included an open-ended question requiring the participants to compare several concepts or summarize a theme, such as "What are the characteristics of vergence movement that distinguish it from saccade and smooth-pursuit?" When answering open-book quizzes, participants could search the videos and the associated discussions. The score for an open-book quiz was evaluated by comparing participants' answers with the standard answers, which consisted of several key points developed by the researchers. The **completeness of the open-book quiz** was calculated by the number of correct points in a participant's answer divided by the number of all the key points in the standard answer. The **completion time of the open-book quiz** was measured by a stopwatch. In addition, the learners' searching behaviors when answering open-book quizzes were recorded, including the number of clicks and the seconds they spent seeking answers by video controllers. In the separate conditions, the platform also recorded the number of times they clicked on the comment list. They were asked to answer an open-book quiz within four minutes.

Workload was measured with a five-item 5-point Likert scale adapted from the National Aeronautics and Space Administration task load index (NASA-TLX) (Hart & Staveland, 1988). It includes mental demand, temporal demand, effort, overall performance, and frustration (Cronbach's alpha = 0.66). The dimension of physical demand in the original scale was excluded because it was less relevant in this experiment.

4.4.5. Post-task interviews about feelings

To ensure that we would not miss any important differences between the two layouts, participants were interviewed about their feelings and preferences and the content of comments.

Although the layout was a between-group variable, at the end of the experiment, the experimenter described the layout that the participants did not see. Participants were asked (a) whether or why they would choose the embedded or separate layout in their learning; (b) how their preference was influenced by the content depth of comments; and (c) their opinions on the effect of the discussions on their learning. Interviews were audio recorded and later transcribed.

4.5. Participants

Forty graduate and undergraduate students at a university in China aged 18 to 27 (M = 21.70, SD = 2.47), including 14 females and 26 males, participated in the experiment. All of them were new to the course, and they were not familiar with neuroscience. They majored in science, technology, engineering, and math (STEM). Because they passed standard college entrance exams in China with similar scores for STEM courses, including biology, we assumed they had similar pre-knowledge and cognitive capabilities.

All of them had experience watching videos with timeline-anchored commenting, such as on Bilibili. Seven of them said they did not like embedded comments and usually hid them when watching a video. Four of those seven participants were assigned to the separate comments group, and three were assigned to the embedded comments group. Other participants decided to hide or show comments depending on the number of comments and the content of the videos and comments.

4.6. Procedure

Each participant took part in the experiment individually in a quiet room. The monitor with the eye tracker was placed on an adjustable desk. A holder was fixed on the edge of the desk in front of the participant. When the eye tracker recorded data, the participant was asked to place their head on the holder to keep it still. The height of the holder and the eye tracker was fixed. We adjusted the height of the desk to suit the height of the participant.

The participants first filled in a background questionnaire and were then asked to complete a practice phase. First, a participant set his/her head on the holder and kept it still. Then, the experimenter adjusted the height of the desk to make it comfortable. The participant then watched a three-minute testing video from the same neuroscience course. After that, the participant moved his/her head away from the holder and completed a testing questionnaire about mind wandering and workload.

Next, the participant began the formal experiment session, that is, watching the two videos with timeline-anchored comments. The order of the videos and the conditions were counterbalanced. Each trial consisted of the following procedures. First, the participant set head on the holder, and the experimenter calibrated the eye tracker and began recording eye-movement data and screen activities. Next, the participant watched one of the videos within a 15-minute timeframe. The participant was told that quizzes would be administered after the video and that he/she could finish watching at any time within the 15 minutes. After watching, the participant moved his/her head away and completed the questionnaire; he/she was asked about the attention they paid to the discussion, mind wandering, and workload. The participant then took a closed-book quiz and an open-book quiz.

After finishing the two learning videos, each participant was briefly interviewed. The entire experiment took approximately 80 minutes. Each participant was paid RMB 100 yuan as remuneration for their time.

4.7. Data analysis

Mixed-model analyses of variance (ANOVAs) were used to determine the main and interaction effects of layout (between-subject variable) and depth (within-subject variable) on performance variables, eye gaze features, and self-reported evaluation of the learning experience. The effect size was measured with a generalized η^2 (Olejnik & Algina, 2003). Regarding post hoc analyses, we conducted pairwise t-tests with false discovery rate (FDR) adjustment and simple effect analyses for significant interaction effects. The clicking behavior data during open-book quizzes were skewed, and therefore pairwise Wilcoxon rank-sum tests were conducted. The transcripts of post-task interviews were analyzed simply by counting participants' layout preference and summarizing the reasons why.

5. Results

Table 2 summarizes how the hypotheses were supported (or otherwise) by the experimental results. The details and the results of the post-task interviews are discussed in the following three subsections.

5.1. Attention allocation to discussion

5.1.1. Attention transitions

The eye-gazing data for two participants were poor in quality, so they were omitted from the analysis. The gazing features were the average values for both eyes. Table 3 indicates the descriptive statistics under various conditions, and Table 4 indicates the hypothesis tests. The layout significantly affected the frequency of glance at the comments ($F_{1,36} = 7.60$, $p < .01$, $\eta^2 = 0.12$) and the average duration of each glance ($F_{1,36} = 19.17$, $p < .01$, $\eta^2 = 0.30$). Learners glanced at the comment AOI more frequently with the embedded layout (M = 0.12, SD = 0.04) than with the separate layout (M = 0.09, SD = 0.05). The average glance duration was shorter under embedded conditions (M = 929 ms, SD = 311 ms) than under separate conditions (M = 1638 ms, SD = 722 ms). Hypothesis 1 was supported.

5.1.2. Gaze fixation and retention of task-oriented comments

The fixation count showed no significant difference under different conditions; Hypothesis 2 was not supported. The main effect of layout on the retention of comments was also not significant; thus, Hypothesis 3a was not supported. However, as illustrated in Table 4 and Figure 4, the interaction of layout and depth significantly affected

Table 2. Summary of hypotheses.

Hypotheses		Results
1. Attention transitions	When learning with the embedded layout, learners switch attention more frequently between the discussion and the lecture.	Supported
2. Gaze fixation on comments	When learning with the embedded layout, learners have more eye gaze fixations on the discussion.	Not supported
3. Retention of comments		
3a. – layout	When learning with the embedded layout, learners memorize more task-oriented comments.	Not supported
3b. – interaction effect	The effect of layout on the retention of comments is greater when the comments are at the surface level than that at the deep level.	Supported
4. Mind wandering	When learning with the separate layout, learners' minds wander more.	Not supported
5. Knowledge retention	When learning with the embedded layout, learners retain more knowledge.	Not supported
6. Knowledge search and integration		
6a. – layout	When learning with the separate layout, learners perform better in knowledge search and integration.	Partially supported by the completion time and clicking behaviors of open-book quizzes
6b. – depth	When learning with deep-level comments, learners perform better in knowledge search and integration.	Partially supported by the completion time of open-book quizzes
6c. – interaction effect	The effect of layout on the performance of knowledge search and integration is stronger when the comments are at a surface level rather than at a deep level.	Partially supported by the completion time and clicking behaviors of open-book quizzes

Table 3. Mean and standard deviation values of attention to discussions and eye-gaze features under four conditions.

	Embedded		Separate	
Variable	Deep *M (SD)*	Surface *M (SD)*	Deep *M (SD)*	Surface *M (SD)*
Attention transitions				
Glance frequency	0.12 (0.03)	0.13 (0.04)	0.10 (0.06)	0.08 (0.04)
Average glance duration [ms]	899 (315)	959 (308)	1,648 (828)	1,628 (611)
Attention allocation				
Fixation count	218 (85)	251 (97)	321 (180)	282 (178)
Retention of task-oriented comments	3.60 (0.82)	4.15 (0.75)	3.65 (1.31)	3.25 (1.37)
Mind wandering				
Mind wandering (post-task self-report)	3.02 (1.09)	3.20 (0.99)	2.85 (0.93)	3.09 (1.01)
Average fixation duration [ms]	279 (84)	303 (66)	262 (92)	294 (60)
Saccade frequency	2.52 (0.40)	2.49 (0.39)	2.70 (0.57)	2.57 (0.44)
Average saccade velocity [° /s]	120 (34)	119 (36)	104 (29)	103 (27)

Table 4. ANOVA summary of main and interaction effects on eye-gaze features (only significant models).

Variable	Effect†	$F_{1,36}$‡	*p*	η^2
Glance frequency	Layout	7.60	< .01	0.12
	Depth	< 0.01	.96	< 0.01
	Interaction	2.95	.09	0.03
Average glance duration	Layout	19.17	< .01	0.30
	Depth	0.08	.79	< 0.01
	Interaction	0.24	.62	< 0.01
Retention of task-oriented comments	Layout	2.46	.13	0.04
	Depth	0.12	.73	< 0.01
	Interaction	4.78	.04	0.05
Average fixation duration	Layout	0.35	.56	< 0.01
	Depth	5.69	.02	0.03
	Interaction	0.11	.74	< 0.01

† "Interaction" means the interaction effect of layout and depth. ‡ for "Retention of task-oriented comments," the degree of freedom of the F value is (1, 38).

the retention of task-oriented comments ($F_{1,38}$ = 4.78, p = .04, η^2 = 0.05). Furthermore, simple effect analysis with FDR adjustment revealed that when task-oriented comments were on the surface level, learners remembered more comments with the embedded layout (M = 4.15, SD = 0.75) than with the separate layout (M = 3.25, SD = 1.37, $F_{1,76}$ = 6.72, p = .02). Pairwise t-tests with FDR adjustment also indicated that the embedded plus surface condition yielded the best recall of task-oriented comments (M = 4.15, SD = 0.75); thus, Hypothesis 3b was supported.

5.2. Mind wandering

No significant difference in self-reported mind wandering was found. The ANOVA models of relevant eye-gazing features indicated no significant effects from layout or depth. With regard to average saccade velocity, a t-test was conducted to compare embedded and separate comment groups without considering the depth. It revealed that learners had higher average saccade velocity with the embedded layout (M = 119.81°/s, SD = 34.60°/s) than with the separate layout (M = 103.53°/s, SD = 27.68°/s, t_{74} = 2.25, p = .03, Cohen's d = 0.52). Learners had a higher saccade velocity with the embedded layout, indicating a lower level of mind wandering (Uzzaman & Joordens, 2011). However, the correlation analysis in this study revealed no significant association between self-reported mind wandering and average saccade velocity or saccade frequency. Hypothesis 4 needs further investigation.

In addition, the average fixation duration was significantly affected by the depth of comments, although the effect size was small ($F_{1,36}$ = 5.69, p = .02, η^2 = 0.03). With a separate layout, surface comments resulted in a longer average fixation

Table 5. Mean and standard deviation values of learning performance and workload under four conditions.

	Embedded		Separate	
Variable	Deep *M (SD)*	Surface *M (SD)*	Deep *M (SD)*	Surface *M (SD)*
Correct rate of closed-book quizzes	0.79 (0.15)	0.78 (0.14)	0.86 (0.12)	0.82 (0.13)
Performance of open-book quizzes				
Completeness	0.54 (0.29)	0.43 (0.20)	0.63 (0.27)	0.42 (0.14)
Completion time [seconds]	218.90 (24.53)	201.55 (45.59)	179.80 (52.09)	195.70 (57.10)
Clicking behaviors of open-book quizzes				
Seeking time [s] -video controller	6.50 (9.76)	9.40 (20.08)	4.89 (14.31)	13.53 (38.89)
Seeking count -video controller	4.35 (4.21)	4.35 (7.41)	1.89 (2.75)	1.94 (1.54)
Seeking count -comment list	-	-	1.42 (1.77)	0.58 (0.96)
Workload	4.39 (0.82)	4.24 (0.72)	4.01 (1.00)	4.12 (0.79)

Table 6. ANOVA summary of main and interaction effects on learning performance (only significant models).

Variable	Effect	$F_{1,38}$	*p*	η^2
Completeness of open-book quizzes	Layout	0.50	.48	< 0.01
	Depth	9.79	< .01	0.11
	Interaction	0.87	.36	0.01
Completion time of open-book quizzes	Layout	3.72	.06	0.06
	Depth	0.01	.94	< 0.01
	Interaction	3.44	.07	0.03

duration (M = 299 ms, SD = 63 ms) than deep comments (M = 271 ms, SD = 88 ms).

5.3. Cognitive learning and workload

5.3.1. Closed-book quizzes

As Tables 5 and 6 indicate, no significant effect on closed-book quiz performance was found; Hypothesis 5 was, therefore, not supported.

5.3.2. Open-book quizzes

Hypothesis 6a was supported by the completion time of the open-book quiz. The effect of layout was marginally significant ($F_{1,38}$ = 3.72, p = .06, η^2 = 0.06). Hypothesis 6b was supported by the completeness of the answers to open-book quizzes. The effect of depth was significant ($F_{1,38}$ = 9.79, p < .01, η^2 = 0.11). Deep-level comments helped learners perform better in open-book quizzes (M = 0.58, SD = 0.28) than surface-level comments (M = 0.43, SD = 0.17). Hypothesis 6 c was also supported by the completion time. Pairwise t-tests of completeness indicated that the separate plus deep condition (M = 0.63, SD = 0.27) yielded a higher score than the embedded plus surface condition (M = 0.43, SD = 0.20, t_{38} = 2.66, p = .04, d = 0.84). The interaction effect of layout and depth on the completion time was marginally significant ($F_{1,38}$ = 3.44, p = .07, η^2 = 0.03). Furthermore, simple effect analysis with FDR adjustment indicated that with deep-level comments, learners completed open-book quizzes more quickly with the separate layout (M = 179.80, SD = 52.09) than with the embedded layout (M = 218.90, SD = 24.53, $F_{1,76}$ = 7.07, p = .02; see Figure 5).

The results of clicking data indicated that separate comments facilitated searching. The clicking log of one participant was not recorded for technical reasons. Pairwise Wilcoxon rank sum tests with FDR adjustment revealed that the participants spent significantly less time seeking information under the separate plus deep condition (M = 4.89, SD = 14.31) than under the two embedded conditions (p values < .05) and the separate plus surface condition (p = .051). The reason was that the separate layout allowed the learners to search the video by clicking comments in the sidebar list, which saved time in

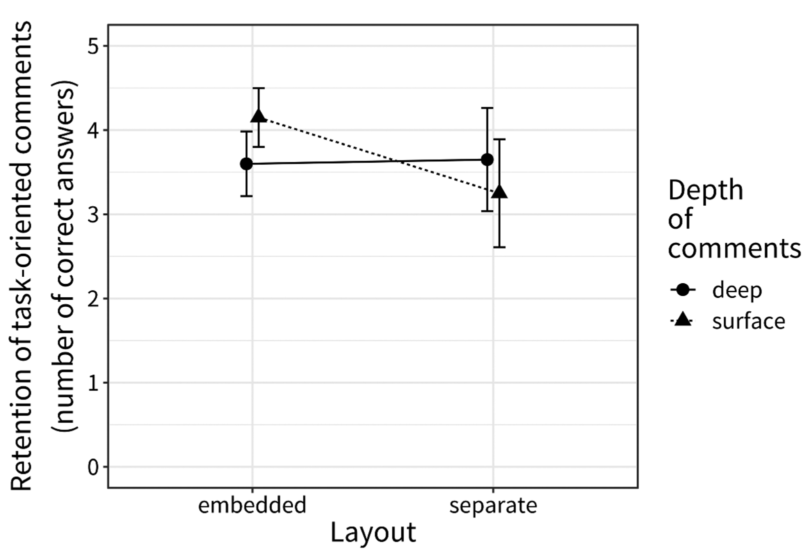


Figure 4. Retention of task-oriented comments (significant interaction effect; N = 40).

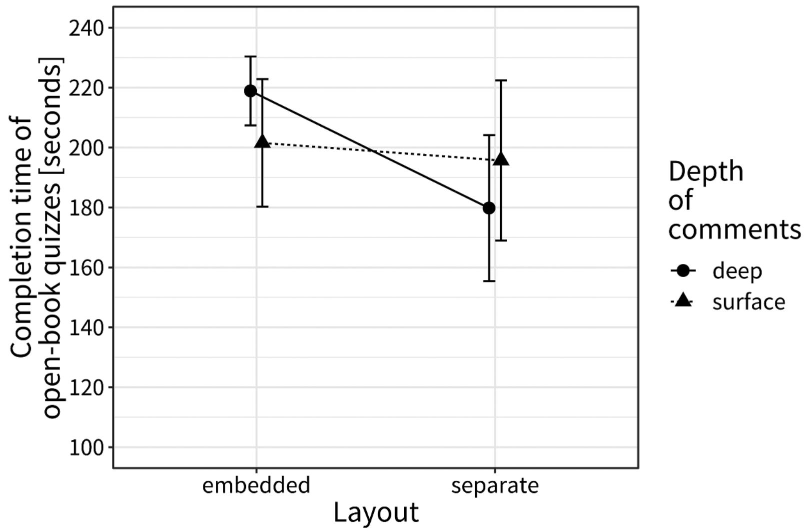


Figure 5. Completion time for open-book quizzes (marginal significant effect of layout and interaction effect, N = 40).

answering open-book quizzes. Additionally, the learners clicked more comments on the sidebar list to answer open-book quizzes when the comments were at the deep level (M = 1.42, SD = 1.77) than when they were at the surface level (M = 0.58, SD = 0.96, p = .07).

5.3.3. Workload

Despite the tendency for the workload to be higher under embedded and deep conditions, no significant difference was found in the study. Nevertheless, it was associated with glance frequency (Pearson's r = 0.31, p < .01). This suggested that more transitions between videos and comments caused higher workload.

5.4. Post-task interviews

The interviews suggested that participants did not prefer one layout over another. Generally, they preferred the layout that they happened to have worked with. After the two layouts were introduced, 14 participants (11 from the embedded group) preferred the embedded layout, and 15 participants (11 from the separate group) preferred the separate layout.

Seven participants mentioned that comments at different depths were suitable for display in different layouts; that is, they experienced the interaction effect of the layout and content depth. They stated that deep-level comments were usually longer or required more time to process, and they should therefore be placed in the permanent separate list, where they could be more easily retrieved, read, and understood. Surface-level comments, such as explanations of concepts in the lecture slides, were better embedded in the video (e.g., in a space next to the concept), so the viewer would not have to move their eyes outside the video area.

6. Discussions

To promote discussion in video-based learning, platforms and previous research have incorporated timeline-anchored comments, which are either embedded in the video or listed beside it. Empirical research has found both cognitive and social benefits of timeline-anchored commenting, but the effect of the visual design of the comments on learners' eye gazes, attention, and learning has not been widely investigated. Therefore, this study applied eye tracking to explore how attention and learning are affected by the layout (embedded or separate) and the depth of the content of comments. In this section, we discuss our key findings, design implications, and limitations.

6.1. Key findings

6.1.1. Attention allocation strategy during video-based learning

With either the embedded or the separate layout, learners switch their attention frequently between the video and timeline-anchored comments, as indicated by the fact that our participants focused on the video for only 10 seconds on average before switching to the discussion, where they spent approximately 1.3 seconds before switching back to the video. Given that the comments in the experiment contained 24 Chinese characters on average and that the average reading speed of Chinese young adults was approximately five to seven Chinese characters per second, as shown (Chan & Lee, 2005; Shu et al., 2011), it can be inferred that learners only skim, rather than read, the comments, to be able to keep up with the video and comments simultaneously.

The results suggest that the participants adopted a concurrent multitasking strategy, i.e., performing both tasks (i.e., watching the video and reading the comments) concurrently (Salvucci et al., 2009), even at the cost of performance quality of individual tasks (e.g., skimming long comments) when watching a lecture video with timeline-anchored comments. This phenomenon indicated that learners may have continuously integrated various elements of multimedia materials to form a mental model of the knowledge, as found in the multimedia learning literature. Such integration requires learners to simultaneously acquire different representations (i.e., video lectures and textual comments), which demands more cognitive resources (Schnotz & Bannert, 2003).

This phenomenon matches well with a number of empirical studies using the eye-tracking methodology, which found significant associations between learners' allocation of visual attention and the level of knowledge integration (Krejtz et al., 2016; Mason et al., 2015; O'Keefe et al., 2014). The majority of previous research, however, investigated the effect of supplementing textual materials with graphic representations (e.g., static pictures, animations, and interactive simulations). Our study examined the effect of accompanying video content with timeline-anchored comments. In this scenario, the media richness is high, and both video and comment contents are time-bound. Continuous and simultaneous integration poses high cognitive demands. Our results showed a tendency of learners to trade off the performance quality of individual tasks for the concurrency of processing.

Comparisons between the two layout conditions showed that the embedded layout promotes more rapid gaze transitions between video and comments than the separate layout. With the embedded layout, the temporal pattern was, on average, 8 seconds on the video and 0.9 seconds on the comments, whereas the pattern with the separate layout was approximately 11 seconds on the video and 1.6 seconds on the comments. This effect can be explained by the two nuisance factors from SEEV theory: the embedded layout reduces the muscular *effort* required to move from gaze to comments, and the comments float on top of the video, resulting in a higher level of *salience*. Both are attributed to the higher attractiveness of comments and encourage more frequent gaze transitions. The more frequent switching of visual attention between the video and comments may contribute to a better memory of the comment content, at least for surface-level discussions, in the embedded layout, which will be discussed in a later section.

Finally, our study found no significant difference in self-reported mind wandering between different conditions, although participants using the embedded layout had a higher saccade velocity, which has been found to be associated with less mind wandering (Uzzaman & Joordens, 2011). We suspected that this might be attributed to the short duration of the learning session (approximately 10 minutes) in our study. Students are more likely to experience mind wandering in longer lectures (Risko et al., 2012; Szpunar, Moulton et al., 2013). For example, in Risko et al.'s study, students' minds wandered much more frequently in the first half of a 60-minute lecture (35%) than in the second half of the lecture (52%). Participants may have controlled and modified their behaviors because of the awareness of being watched in the laboratory. In addition, most previous research on mind wandering in a class focused on off-task multitasking (e.g., Jamet et al., 2020; Loh et al., 2016), whereas the focus of this study was on-task multitasking, i.e., watching the lecture video and reading the discussion texts simultaneously. Therefore, future research is needed to examine how these discussions related to video lectures affect mind wandering in class.

6.1.2. Retention of timeline-anchored comments

As discussed above, the more frequent gaze transitions may contribute to a better memory of surface-level task-oriented comments with the embedded layout than with the separate layout. Previous research has found that more gaze transitions are associated with more integration of information from multiple sources (Krejtz et al., 2016; Mason et al., 2015; O'Keefe et al., 2014). This finding is also in accordance with multimedia learning theory (Ginns, 2006; Mayer, 2009), which posits that a higher spatial contiguity among multimedia learning materials leads to greater retention of the materials.

However, no positive effect of the embedded layout on retention was shown for deep-level comments. This resulted from the conflict between the layout and the content. On the one hand, the embedded layout presents comments in transient moving texts, which need more time to read and catch up on, potentially overwhelming learners (De Koning et al., 2010; Leahy & Sweller, 2011; Sweller, 2011; Uetsuki et al., 2017). On the other hand, deep comments required more time and cognitive effort to process. Due to this conflict, the participants may have given up processing deep comments for the sake of keeping up with the video. This supposition was supported by our eye-tracking data; as an indicator of the difficulty or depth of cognitive processing (Engbert et al., 2005; Nuthmann et al., 2010), the duration of the time focused on deep comments was not longer, but was actually slightly shorter, than that on surface comments. This is consistent with the attention management strategy of learning such videos that we previously identified: catching up on the video content is the top priority and additional information will be ignored if it is difficult to follow or process.

6.1.3. Learning performance and workload

Although the embedded layout improves participants' memory of surface-level comments, it does not translate into improvement in learning outcomes in the quizzes. This might be attributed to the fact that the knowledge mentioned in the surface-level comments was not crucial to or was not tested in the quizzes in this study by chance. Future investigations are needed to test the effect of surface-level comments with different videos and comments.

Deep-level comments, although they tend to be inadequately processed during video learning, lead to better performance in the open-book test than surface-level comments. When deep-level comments are provided in a separate layout, learners complete the knowledge search integration task more quickly. The open-book questions were designed to require the participants to compare and integrate multiple concepts or knowledge points from the video they had studied. They examined how well the tool supports learners in navigating, searching, and integrating knowledge when they revisit a video rather than while watching the video. Previous research found that such tasks can be facilitated by visual cues that provide a map or overview of the learning materials (Hsu & Schwen, 2003; Li et al., 2013; Lu et al., 2018). Our results suggest that deep-level comments may play the role of such cues, i.e., provide an overview of key points of the learning content, particularly when they are presented in a separate layout. They help learners quickly locate the relevant time point in the video and access the related video content.

6.2. Implications

A major implication for designers of timeline-anchored commenting systems is that different layouts serve different outcomes with different types of comments. The embedded layout encourages more frequent attention transition, enhances knowledge retention of surface-level comments that require less cognitive effort (e.g., simple explanations, repetition for the purpose of emphasis, and translations), and has the potential to reduce mind wandering during learning, as indicated by the higher saccade velocity. Therefore, the embedded layout should be deployed when the majority of comments are surface-level and when the purpose is to actively engage learners during the video session. In contrast, the separate layout is more appropriate in the case of deep-level comments and when the expected usage is to facilitate the search and integration of related information after video viewing.

For scenarios in which a variety of discussion depths and expected outcomes may coexist, learners can be provided controls to select the proper layout based on their preference and specific scenario requirements. In addition, the system may help to identify the cognitive depth of accumulated timeline-anchored comments through content analysis and adapt the layout accordingly. Recent researchers (Joksimović et al., 2019) have proposed coding instruments to identify the depth of self-reflection that is revealed in annotations to video lecture timelines. Such text-analytic methods can be adopted to design platforms that can adapt the layout to content depth automatically.

Another approach to utilize timeline-anchored comments is to extract labels or keywords for overview and searching. Recently, researchers have developed methods for extracting and visualizing the keywords of a video by analyzing timeline-anchored comments and scripts (Lu et al., 2018; Zhu et al., 2020), as well as learners' emotional and feeling states (Chatti et al., 2016; Peng et al., 2019; Sung et al., 2016, 2017). Our results indicate that extracted keywords, which can be considered concise deep comments, can be displayed in a separate manner, either in the sidebar list or along the play timeline, to facilitate reviewing, searching, and comprehending relevant knowledge.

6.3. Limitations

First, limitations of the sample need to be acknowledged. To avoid confounding influences brought by different academic disciplines of learners and different levels of familiarity with topics, we recruited only STEM students with little knowledge about the learning materials in the experiment. Whether the results can be applied to students of other majors or students who are knowledgeable about the topic requires further investigation. In addition, video-based learning has become quite a ubiquitous way to learn, not only in formal approaches (e.g., MOOCs or online learning materials provided by school teachers) but also informally (e.g., from YouTube videos). The latter approach implies an even more diversified population of learners, which may be neither young nor well educated. Generalizing findings from the current study to informal learning contexts requires further empirical investigations. Second, the experiment was conducted in the laboratory, and the participants studied two short lecture videos in a simulated environment. The participants are less likely to experience mind wandering in this setting than when they watch a video at home. In addition, mind wandering was measured by an after-task survey, which may lose the freshness of data. To better examine the influence of the layout and depth of comments on mind wandering, future research could adopt real-world learning environments, longer learning tasks, and more immediate measurement of mind wandering, e.g., probe methods or psychophysiological measurement.

7. Conclusion

This study investigated how attention, eye gazing, and learning during lecture videos were affected by the layout and the depth of timeline-anchored comments. The results indicated that timeline-anchored comments led to frequent attentional transitions between the video and the comments. With the embedded layout, which displays comments as texts on videos, learners switched their attention more frequently between the comments and the videos and recalled more surface-level comments. The separate layout, which displayed comments in a sidebar list, facilitated searching and overviews of relevant knowledge and increased performance in open-book quizzes, especially when the comments were at a deep level. We also considered the implications of better presented timeline-anchored comments in the future design of video-based learning.

Note

1. https://v1-www.xuetangx.com/courses/course-v1:NTHU+MOOC_02_002+2018_T1/about?

Funding

This study was supported by the National Natural Science Foundation of China (Project no. 71942005).

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*(5), 427–432. https://doi.org/10.1111/1467-9280.01458

Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250. https://doi.org/10.1111/j.1467-8535.2009.01029.x

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). Cambridge University Press.

Baker, C. (2010). The impact of instructor immediacy and presence for online student affective learning, cognition, and motivation. *The Journal of Educators Online*, *7*(1), 1–30. https://doi.org/10.9743/JEO.2010.1.2

Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: Spatial contiguity effects as trade-off between gaze switches and memory load. *Applied Cognitive Psychology*, *26*(6), 863–871. https://doi.org/10.1002/acp.2887

Bloom, B. S. (1956). *Taxonomy of educational objectives. Vol. 1: Cognitive domain*. McKay.

Brooker, A., Corrin, L., De Barba, P., Lodge, J., & Kennedy, G. (2018). A tale of two MOOCs: How student motivation and participation predict learning outcomes in different MOOCs. *Australasian Journal of Educational Technology*, *34*(1), 73–87. https://doi.org/10.14742/ajet.3237

Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2018). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology*, *88*(1), 118–137. https://doi.org/10.1111/bjep.12181

Cegala, D. J. (1981). Interaction involvement: A cognitive dimension of communicative competence. *Communication Education*, *30*(2), 109–121. https://doi.org/10.1080/03634528109378460

Chan, A. H. S., & Lee, P. S. K. (2005). Effect of display factors on Chinese reading times, comprehension scores and preferences. *Behaviour & Information Technology*, *24*(2), 81–91. https://doi.org/10.1080/0144929042000267073

Chatti, M. A., Marinov, M., Sabov, O., Laksono, R., Sofyan, Z., Yousef, A. M. F., & Schroeder, U. (2016). Video annotation and analytics in CourseMapper. *Smart Learning Environments*, *3*(1), 10. https://doi.org/10.1186/s40561-016-0035-1

Chávez, J., Montaño, R., & Barrera, R. (2016). Structure and content of messages in an online environment: An approach from participation. *Computers in Human Behavior*, *54*, 560–568. https://doi.org/10.1016/j.chb.2015.08.046

Chen, Y., Gao, Q., & Rau, P.-L. P. (2017). Watching a movie alone yet together: Understanding reasons for watching danmaku videos. *International Journal of Human–Computer Interaction*, *33*(9), 731–743. https://doi.org/10.1080/10447318.2017.1282187

Chen, Y., Gao, Q., Yuan, Q., & Tang, Y. (2019). Facilitating students' interaction in MOOCs through timeline-anchored discussion. *International Journal of Human–Computer Interaction 35*(19) , 1781–1799. https://doi.org/10.1080/10447318.2019.1574056

Chiu, K. F. T., & Hew, K. F. T. (2018). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian Journal of Educational Technology*, *34*(4), 16–28. https://doi.org/10.14742/ajet.3240

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005

De Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education*, *55*(2), 681–691. https://doi.org/10.1016/j.compedu.2010.02.027

Dinsmore, D. L., & Alexander, P. A. (2016). A multidimensional investigation of deep-level and surface-level processing. *The Journal of Experimental Education*, *84*(2), 213–244. https://doi.org/10.1080/00220973.2014.979126

Dolmans, D. H., Loyens, S. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: A review of the literature. *Advances in Health Sciences Education*, *21*(5), 1087–1112. https://doi.org/10.1007/s10459-015-9645-6

Eilertsen, T. V., & Valdermo, O. (2000). Open-Book assessment: A contribution to improved learning? *Studies in Educational Evaluation*, *26*(2), 91–103. https://doi.org/10.1016/S0191-491X(00)00010-9

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777. https://doi.org/10.1037/0033-295X.112.4.777

Garrison, D. R., & Arbaugh, J. B. (2007). Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education*, *10*(3), 157–172. https://doi.org/10.1016/j.iheduc.2007.04.001

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

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, *28*(2), 115–152. https://doi.org/10.1023/A:1003764722829

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Henri, F. (1992). Computer conferencing and content analysis. *Collaborative learning through computer conferencing* (pp. 117–136). Springer.

Hsu, Y., & Schwen, T. M. (2003). The effects of structural cues from multiple metaphors on computer users' information search performance. *International Journal of Human-Computer Studies*, *58*(1), 39–55. https://doi.org/10.1016/S1071-5819(02)00125-8

Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human Behavior*, *106*, 106264. https://doi.org/10.1016/j.chb.2020.106264

Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, *32*, 47–53. https://doi.org/10.1016/j.chb.2013.11.013

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. https://doi.org/10.1037/a0026923

Johnson, D. (2013). Polyphonic/Pseudo-synchronic: Animated writing in the comment feed of Nicovideo. *Japanese Studies*, *33*(3), 297–313. https://doi.org/10.1080/10371397.2013.859982

Joksimović, S., Dowell, N., Gašević, D., Mirriahi, N., Dawson, S., & Graesser, A. C. (2019). Linguistic characteristics of reflective states in video annotations under different instructional conditions. *Computers in Human Behavior*, *96*, 211–222. https://doi.org/10.1016/j.chb.2018.03.003

Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCs). *Computers & Education*, *122*, 9–22. https://doi.org/10.1016/j.compedu.2018.02.013

Ke, F., & Kwak, D. (2013). Online learning across ethnicity and age: A study on learning interaction participation, perception, and learning satisfaction. *Computers & Education*, *61*, 43–51. https://doi.org/10.1016/j.compedu.2012.09.003

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Krejtz, K., Duchowski, A. T., Krejtz, I., Kopacz, A., & Chrząstowski-Wachtel, P. (2016 Gaze transitions when learning with multimedia). *Journal of Eye Movement Research 9*(1), 1–17. https://doi.org/10.16910/jemr.9.1.5

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*(6), 943–951. https://doi.org/10.1002/acp.1787

Lee, Y.-C., Lin, W.-C., Cherng, F.-Y., Wang, H.-C., Sung, C.-Y., & King, J.-T. (2015). Using time-anchored peer comments to enhance social interaction in online educational videos. *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 689–698). https://doi.org/10.1145/2702123.2702349

Leng, J., Zhu, J., Wang, X., & Gu, X. (2016). Identifying the potential of Danmaku video from eye gaze data. *Advanced Learning Technologies (ICALT), 2016 IEEE 16th international conference on* (pp. 288–292). http://ieeexplore.ieee.org/abstract/document/7756979/

Li, C., & Lalani, F. (2020, April 29). *The COVID-19 pandemic has changed education forever. This is how*. World Economic Forum. https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/

Li, L.-Y., Chen, G.-D., & Yang, S.-J. (2013). Construction of cognitive maps to improve e-book reading and navigation. *Computers & Education*, *60* (1), 32–39. https://doi.org/10.1016/j.compedu.2012.07.010

Lin, X., Huang, M., & Cordie, L. (2018). An exploratory study: Using Danmaku in online video-based lectures. *Educational Media International*, *55*(3), 273–286. https://doi.org/10.1080/09523987.2018.1512447

Lindquist, S. I., & McLean, J. P. (2011). Daydreaming and its correlates in an educational environment. *Learning and Individual Differences*, *21* (2), 158–167. https://doi.org/10.1016/j.lindif.2010.12.006

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *44*(1), 87. https://doi.org/10.1037/h0084237

Loh, K. K., Tan, B. Z. H., & Lim, S. W. H. (2016). Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Computers in Human Behavior*, *63*, 943–947. https://doi.org/10.1016/j.chb.2016.06.030

Lu, Z., Heo, S., & Wigdor, D. J. (2018). StreamWiki: Enabling viewers of knowledge sharing live streams to collaboratively generate archival documentation for effective in-stream and Post Hoc learning. *Proceedings of the ACM on human-computer interaction, 2 (CSCW)* (pp. 112).

Ma, X., & Cao, N. (2017). Video-based evanescent, anonymous, asynchronous social interaction: Motivation and adaption to medium. *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 770–782).

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect. *Learning and Instruction*, *61*, 23–34. https://doi.org/10.1016/j.learninstruc.2018.12.001

Mason, L., Pluchino, P., & Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly*, *48*(2), 199–214. https://doi.org/10.1002/rrq.41

Mason, L., Tornatora, M. C., & Pluchino, P. (2015). Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: An eye-movement study. *Reading and Writing*, *28*(6), 851–872. https://doi.org/10.1007/s11145-015-9552-5

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press New York.

Mcmillan, R., Kaufman, S., & Singer, L. J. (2013). Ode to positive constructive daydreaming. *Frontiers in Psychology*, *4*, 626. https://doi.org/10.3389/fpsyg.2013.00626

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin*, *136*(2), 188–197. https://doi.org/10.1037/a0018298

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382. https://doi.org/10.1037/a0018924

O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., & Plass, J. L. (2014). Learning from multiple representations: An examination of fixation patterns in a science simulation. *Computers in Human Behavior*, *35*, 234–242. https://doi.org/10.1016/j.chb.2014.02.040

Offir, B., Lev, Y., & Bezalel, R. (2008). Surface and deep learning processes in distance education: Synchronous versus asynchronous systems. *Computers & Education*, *51*(3), 1172–1183. https://doi.org/10.1016/j.compedu.2007.10.009

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434. https://doi.org/10.1037/1082-989X.8.4.434

Pashler, H. (1988). Cross-dimensional interaction and texture segregation. *Perception & Psychophysics*, *43*(4), 307–318. https://doi.org/10.3758/BF03208800

Peng, Q., Qie, N., Yuan, L., Chen, Y., & Gao, Q. (2019). Design of an online education evaluation system based on multimodal data of learners. *International conference on cross-cultural design*. In press.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*(9), 1300–1310. https://doi.org/10.1177/0956797610378686

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology*, *26*(2), 234–242. https://doi.org/10.1002/acp.1814

Rovai, A. P. (2002). Sense of community, perceived cognitive learning, and persistence in asynchronous learning networks. *The Internet and Higher Education*, *5*(4), 319–332. https://doi.org/10.1016/S1096-7516(02)00130-6

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1819–1828).

Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, *36*, 11–26. https://doi.org/10.1016/j.learninstruc.2014.11.002

Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In Was, C., Sansosti, F., and Morris, B. (Eds.). *Eye-tracking technology applications in educational research* (pp. 143–164). IGI Global.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*(2), 100–110. https://doi.org/10.1016/j.learninstruc.2009.02.011

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13* (2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning out while reading: Evidence for dissociations between experience and metaconsciousness. In Levin, D. T. (Ed.). *Thinking and seeing: Visual metacognition in adults and children* (pp. 203). The MIT Press.

Shin, H., Ko, E.-Y., Williams, J. J., & Kim, J. (2018). Understanding the effect of In-Video prompting on learners and instructors. *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 319).

Shu, H., Zhou, W., Yan, M., & Kliegl, R. (2011). Font size modulates saccade-target selection in Chinese reading. *Attention, Perception, & Psychophysics*, *73*(2), 482–490. https://doi.org/10.3758/s13414-010-0029-y

Singh, A.-M., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology*, *26*(6), 848–853. https://doi.org/10.1002/acp.2885

Sins, P. H., Van Joolingen, W. R., Savelsbergh, E. R., & Van Hout-Wolters, B. (2008). Motivation and performance within a collaborative computer-based modeling task: Relations between students' achievement goal orientation, self-efficacy, cognitive processing, and achievement. *Contemporary Educational Psychology*, *33*(1), 58–77. https://doi.org/10.1016/j.cedpsych.2006.12.004

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*(6), 1144–1150. https://doi.org/10.3758/MC.36.6.1144

Smallwood, J. (2010). Why the global availability of mind wandering necessitates resource competition: Reply to McVay and Kane. *Psychological Bulletin*, *136*(2), 202–207. https://doi.org/10.1037/a0018673

Sorden, S. D. (2005). A cognitive approach to instructional design for multimedia learning. *Informing Science Journal*, *8*, 263–279. https://doi.org/10.28945/498

Sung, C.-Y., Huang, X.-Y., Shen, Y., Cherng, F.-Y., Lin, W.-C., & Wang, H.-C. (2016). ToPIN: A visual analysis tool for time-anchored comments in online educational videos. *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems* (pp. 2185–2191). https://doi.org/10.1145/2851581.2892327

Sung, C.-Y., Huang, X.-Y., Shen, Y., Cherng, F.-Y., Lin, W.-C., & Wang, H.-C. (2017). Exploring online learners' interactive dynamics by visually analyzing their time-anchored comments. *Computer Graphics Forum*, *36*(7), 145–155. https://doi.org/10.1111/cgf.13280

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: From the classroom to online learning. *Frontiers in Psychology*, *4*, 1–7. https://doi.org/10.3389/fpsyg.2013.00495

Uetsuki, M., Watanabe, J., Ando, H., & Maruya, K. (2017). Reading traits for dynamically presented texts: Comparison of the optimum reading rates of dynamic text presentation and the reading rates of static text presentation. *Frontiers in Psychology*, *8*, 1–10. https://doi.org/10.3389/fpsyg.2017.01390

Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, *20*(4), 1882–1886. https://doi.org/10.1016/j.concog.2011.09.010

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, *87*(2), 246. https://doi.org/10.1037/0022-3514.87.2.246

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, *16*(4), 359–384. https://doi.org/10.1007/s10648-004-0005-y

Weinstein, Y. (2017). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior research methods 50*(2), 642–661. https://doi.org/10.3758/s13428-017-0891-9

Wickens, C. D., & McCarley, J. S. (2007). *Applied attention theory*. CRC Press.

Wise, A. F., & Cui, Y. (2018). Unpacking the relationship between discussion forum participation and learning in MOOCs: Content is key. *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 330–339).

Yao, Y., Bort, J., & Huang, Y. (2017). Understanding Danmaku's potential in online video learning. *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems* (pp. 3034–3040). https://doi.org/10.1145/3027063.3053258

Yardi, S. (2006). The role of the backchannel in collaborative learning environments. The International Conference of the Learning Sciences. Bloomington, Indiana, USA. *2*, 852–858. https://dl.acm.org/doi/10.5555/1150034.1150158

Yousef, A. M. F., Chatti, M. A., Schroeder, U., & Wosnitza, M. (2015). A usability evaluation of a blended MOOC environment: An experimental case study. *The International Review of Research in Open and Distributed Learning*, *16*(2), 69–93. https://doi.org/10.19173/irrodl.v16i2.2032

Zhu, W., Zang, J., & Tobita, H. (2020). Wordy: Interactive word cloud to summarize and browse online videos to enhance eLearning. *2020 IEEE/SICE international Symposium on System Integration (SII)* (pp. 879–884). https://doi.org/10.1109/SII46433.2020.9026306

About the Authors

Yue Chen is a postdoctoral researcher in the Department of Industrial Engineering at Tsinghua University, Beijing, China. She received her PhD from Tsinghua University. Her research interests include user-centered design, human-computer interaction, social media, and computer assisted learning.

Qin Gao is currently an associate professor in the Department of Industrial Engineering at Tsinghua University, Beijing, China. She received her PhD from Tsinghua University. Her primary research interests include user-centered design, cognitive ergonomics, human-computer interaction, and decision-making.

Ge Gao is an assistant professor at the College of Information Studies (iSchool), the University of Maryland, College Park. She obtained her PhD at Cornell University. Her research focuses on computer-supported cooperative work, human-centered computing, and global teams.