iasdr 2021 WITH DESIGN
PolyU Design School of Design The Hong Kong Polytechnic University 5–9 December 2021

Interface, Pedagogical Agents and Chatbox in Virtual Learning Environments: An Eye Tracking Experiment

Wei, Lai[a]; Chow, Kenny K. N.[b]

[a] School of Design, The Hong Kong Polytechnic University, Hong Kong, China School of Design, The Hong Kong Polytechnic University, Hong Kong, China

[b] School of Design, The Hong Kong Polytechnic University, Hong Kong, China School of Design, The Hong Kong Polytechnic University, Hong Kong, China

* sdknchow@polyu.edu.hk

Scholars in education over the past few decades have discussed at length computer-based interactive learning through the conceptual framework of cognitive science. Yet, very few studies look at the design of visual content in technology-mediated learning and its effects on information dissemination. During the COVID-19 pandemic, the situation in online education revealed that the potential of visual and non-verbal design attributes in learning is underexplored. This study employed a 2X2 experimental design. An Eyelink 1000 plus eye-tracker was used to record student participants' eye movements to determine how they looked at the interface layout, the chat box, and pedagogical agents with text cues. The participants' learning performance were measured with retention and transfer tests. Results showed that the interface design, chat box and pedagogical agents interfere with students' learning performance and visual attendance, which proves spatial contiguity effect, coherence effect and redundancy effect from the cognitive theory of multimedia learning. The present study results support the hypothesis that visual design is fundamental for effective virtual learning environments.

Keywords: Virtual Learning Environments; Learning Information Transmission; Multimedia Learning; Interface Design; Pedagogical Agents; Chatbox; Eye Tracking;

1 Introduction

Since the COVID-19 pandemic disrupted regular global mobility, industries had shift away from conventional communication. Face-to-face education abruptly shifted to an online learning model (Russ & Hamidi 2021; Fakhrunisa & Prabawanto 2020; Gillis & Krull 2020; Wang & Cruz 2020; Zeng 2020). Existing online learning platforms include visual elements in virtual learning environments (VLEs) such as presented materials, lecturers and interaction. Presentations consist of slides with text, pictures and videos. Lecturers are human or virtual agents who appear in a virtual classroom and narrate educational content. Interactive elements facilitate communication with lecturers and learning peers. Educational institutions and communication technology companies (e.g., Zoom,

Team and Tencent) utilize the visual elements to design VLEs for improving students' learning engagement and performance.

According to the cognition theory of multimedia learning (CTML; Mayer 2002) and cognitive load theory (CLT; Sweller 1994), integrated and coherent information stimulates students' dual-channel (verbal and pictorial modalities) in working memory for reducing cognitive overload and promoting the transferral of novel knowledge into long-term memory during the learning process. However, the widely-used VLE interface is designed with distinctly dispersed regions and monotonous information, and students have to divide attention to acquire information. Early quantitative studies support that visual elements (e.g., virtual agents, social cues and interaction function) in VLEs positively affect learning performance (Zeng 2020; Hung & Chen 2018). There is a need for further research evaluating the usability of a VLE interface layout and its' effect on learners' learning performance and experience.

Many studies explored the relationship between learning performance and visual attendance, and the results indicated the longer the eye fixation duration, the higher the learning performance of the students (Stull, et al. 2018; Rey 2014). In addition, integrated visual information in VLEs increases visual attendance over that of dispersed information in VLEs (Wang, et al. 2018). However, using eye tracking to investigate the particular effect of three visual elements in VLEs on visual attendance is underexplored. This study evaluates the effect of interface layout, pedagogical agents and a chat box on learning performance and the duration of learners' eye fixation.

2 Relevant Literature

2.1 Interface Design

In previous literature, research on learning information transmission pays great attention to presented materials and lecturers' performance (Rapanta, et al. 2020; Stull, et al. 2018; Schroeder, et al. 2013). From the perspective of the usability of the VLE interface layout, studies investigating the effect of the distribution of the VLE interface on learning process had rarely been proposed. Existing VLE interfaces present a large region containing teaching materials, and normally the right-side of the interface contains an interactive region for students to communicate with lecturers and peers. The design of this interface aims to provide a teaching materials region and a separate interactive region to assist learners in following the lecture (Farhan, et al. 2019; Hu, et al. 2019; Anon 2018; Turumogon & Baharum 2018; Mirsarraf, et al. 2017). From the perspective of user experience design, the teaching region dominates the largest proportion of the area since it contains the main source of learning information (Hung et al. 2018; Wang, et al. 2018; Shin & Downing 2011). We postulate that the teaching region conveys more information than the interactive region in the VLE interface, and the interactive region essentially performs as a social factor affecting the learning performance and experience (Lam et al. 2019) indicated in social learning theory (SLT; Bandura 1977). To date, few studies have analysed learners' visual attendance through instruments like eye trackers, the duration of learners' fixation on a multi-region interface layout, and the causal effect on learning outcomes.

2.2 Pedagogical Agents

A number of studies point out that lecturers' appearances and facial expressions (Moreno & Mayer 2004; Sutton & Wheatley 2003) and their coherent deictic gestures (Beege, et al. 2020; Davis & Vincent 2019; Wang & Antonenko 2017) all significantly impact students' learning process. In addition, some researchers clarify that both human lecturers and pedagogical agents in VLEs perform almost identical teaching guidance. Learners can perceive the pedagogical agents in the role of lecturers and identify the emotions and gestures from the pedagogical agents (Horovitz & Mayer 2021; Lawson et al. 2021). Pedagogical agents in place of human lecturers' disseminating information in the form of pictures, speaking and animation have proved to be a multimodality of learning information transmission (Castro-Alonso, et al. 2021; Li, et al. 2019; Makransky, et al. 2019; Wang, et al. 2018; Dinçer & Doğanay 2017). Cognitive theory of multimedia learning (CTML; Mayer 2002) theoretically underpins that better transferral of learning information through different modalities, and pedagogical agents in VLEs are a modality that intentionally uses multimedia. To process meaningful learning, some studies found that pedagogical agents used visual cues in VLEs (e.g. highlighting information; Wang et al, 2020; Xie, et al. 2019) facilitate information transmission. Early literature postulates that pedagogical agents lead a learning information-based conversation with learners, which is assumed to affect the learners' learning performance (Schroeder et al. 2013).

2.3 Chatbox

A chat box is a visually represented dialogue with virtual agents serving as dialogue initiators. A chat box is designed to assist learners by outlining the crucial information and enabling a sense of social presence and social identity (Gnewuch 2017; Moreno & Mayer 2004; Biocca 2003; Tu 2000; Short, Williams & Christie 1976). Several studies on text communication (e.g. barrage) examine the effect on viewers' empathy, and the results show that text communication has a significant impact on viewers who need a sense of social presence (Chen, et al. 2019). But the effect of a chat box on learners' visual attention and learning performance is assumed to be negative, due to the fact that information in the chat box is separated from the teaching region. Based on the spatial contiguity effect and the coherence effect proposed in CTML (Mayer 2002), the scattering of information between pedagogical agents and peers is assumed to cause learners' cognitive overload. The visual content of the chat box repeats the key information in the form of text and affects information organization supposed to improve the retention and transfer of knowledge, according to the redundancy effect in CTML (Mayer 2002) and CLT (Sweller 2011). Some studies also pointed out that presenting excessive information would aggravate learners' cognitive load. However, Koning's research demonstrates that whether information is dispersed has no significant effect on increased cognitive load (Koning et al. 2020). There has been little investigation on how a chat box impacts on learning performance. And retention and transfer knowledge is rarely elaborated.

3 Purposes and Hypotheses

The majority of VLE interface layouts use separate regions that are concurrently presented. Based on CTML (Mayer 2002), separated-region interface layouts in VLEs distract students' attention and affect learning performance. On the other hand, pedagogical agents and a chat box in VLEs are proposed to promote learning performance and experience (Moreno & Mayer 2004). The study aimed to investigate the effect of a multi-region interface layout, a chat box and pedagogical agents on students' learning performance and visual attendance. Two research questions were explored. First, how do pedagogical agents and a chat box providing text cues affect students' learning performance and visual attendance? Second, how do VLEs with multiple regions affect students' learning performance and visual attendance? To answer the first question, two hypotheses are stated for evaluating the effect of virtual pedagogical agents (H1 a & b). The second question is designed to examine the effects of a VLE interface layout on learning performance and visual attendance, and two hypotheses are proposed (H2 a & b).

H1 a: Students who learn from the VLEs with pedagogical agents and a chat box will perform better on tests of learning performance than students who learn from the VLEs without pedagogical agents and a chat box.

H1 b: Students who learn from the VLEs with pedagogical agents and a chat box will visually spend more quality time than students who learn from the VLEs without pedagogical agents and chat box.

H2 a: Students who learn from VLEs with multiple regions will perform better on learning outcomes than students who learn from VLEs with a single region.

H2 b: Students who learn from VLEs with multiple regions will visually spend more quality time than students who learn from VLEs' interface with a single region.

4 Materials and Methods

4.1 Experiment Design

Four conditions are designed in the 2X2 experimental design study. In the control condition, the video provides a single region with presented material to group A-1. In the experiment conditions, visual cues manifested by pedagogical agents and a chat box were added for groups A-2, B-1 and B-2. The educational information in text form was transmitted through virtual peers in a chat box in A-2 and B-2 and through a static pedagogical agent in B-1 and B-2. Each video lasted for 5 minutes (Figure 2). Thus, in A-1, the video was presented on a single region in the VLE, and educational information was delivered through PowerPoint slides with text and images. In B-1, the VLE consisted of a single region with PowerPoints containing text, images and a pedagogical agent. Educational information was delivered through two separate regions in A-2, one with PowerPoint slides with text and images, and the other with a chat box for communication with virtual peers. In B-2, the teaching video had two separate regions in the VLE that are PowerPoint slides with text and images

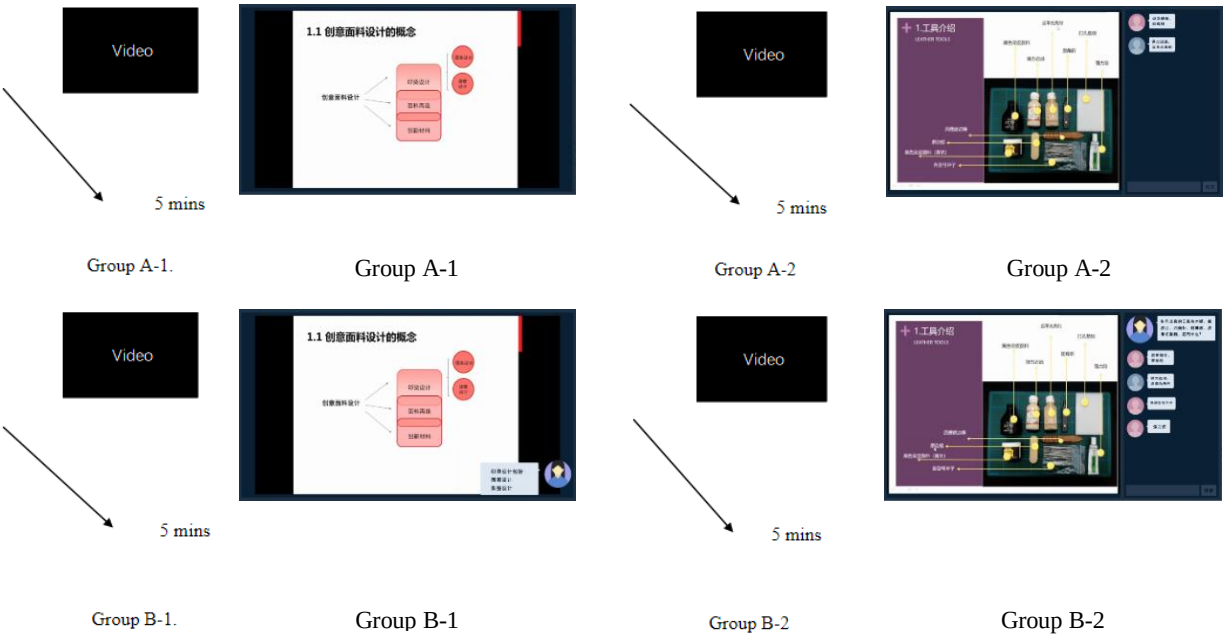

Figure 2. Flow Chart of Experiment and Experiment Design.

dominating the main region, and a secondary interactive region containing a static pedagogical agent and virtual peers in the chat box (Table 1). Two different types of educational narrated videos were utilized, and the two types of knowledge were pretested to show groups had similar levels prior to the experiment. The video narrations for groups A 1 and 2 were about textile design and the videos for groups B-1 and 2 were about leather design. The voice-over presenter in the four videos had been a lecturer at Southwest University for four years. To maintain the learning vitality of students, they are required to finish two five-minute videos. In the videos, PowerPoints were used to display visual content including text, images and instructional videos, verbal source performs accordingly.

Table 1 Condition Explanations of the Experiment Design

Conditions		Explanations
Control Condition	Single-region (A-1) Condition	Video with single-region with teaching materials. Teaching materials dominate main region of interface.
Experimental Conditions	Single-region and Pedagogical Agent (B-1) Condition	Video with single-region. Teaching materials dominate main region of interface, and a static pedagogical agent with textual cues on the bottom right corner of the screen. When teaching video presents learning information, textual cues provided from the pedagogical agent pop up accordingly.
	Multi-region Interface and Chatbox (A-2) Condition	Video with multiple regions. Teaching materials and an interaction region for communicating with virtual peer on the right side of screen.
	Multi-region Interface, Pedagogical Agent and Chatbox (B-2) Condition	Video with multiple regions. Teaching materials dominate main region of interface, a static pedagogical agent with textual cues on the above right corner of the screen and an interaction region for communicating with virtual peer on the right side of screen. When teaching video presents learning information, textual cues provided from the pedagogical agent pop up accordingly.

4.2 Participants

Participants in this study were 60 undergraduate students (17 males and 43 females) from Southwest University in China, and the data from 55 participants was valid. Before the experiment, the participants were evaluated for their subject-related prior knowledge and none of the participants demonstrated high prior knowledge of the materials (mean=8.23, high score=15, SD=3.66). Their ages ranged between 18 and 25 years old, and the mean was 20.4 years old (SD=1.91). All the participants were native Chinese speakers without vision impairment. The participants were randomly assigned to the four conditions. In group A-1, 27 students were assigned as the control group, without pedagogical agents and a chat box. For group B-1, a total of 27 students were assigned to the experimental condition which included a pedagogical agent. In condition A-2, 28 students were assigned to the experimental condition without pedagogical agents but with a chat box. For condition B-2, a total of 28 students were assigned to the experimental condition which included a pedagogical agent and a chat box.

4.3 Devices

An eye tracker, Eyelink 1000 plus, with a sampling rate of 120Hz was used to track eye movements while each participant watched the educational video. As Eyelink 1000 plus is a remote eye tracker, each participant could sit about 50 cm in front of the eye tracker while having their heads positioned a natural distance from the screen. Two 15-inch laptop were used, one for monitoring the eye movements of participants, and another one for displaying the educational videos. An independent room was set up with a white desk and participants faced a white wall during the experiment.

4.4 Measures of Demographic, Prior Knowledge and Learning Experience

The questionnaires were designed based on Mayer's multimedia learning study and related literature (Wang, et al. 2020; Wang, et al. 2018; Rey 2014; Mayer 2002) and collected participants' demographic information including gender, age, major and grade of college education. The prior knowledge questionnaire was employed to examine the participants' knowledge about the educational videos, which is required to take as a relative consideration of their learning performance. A learning experience questionnaire was designed based on previous related literature (Li et al. 2019), and it evaluates learning experience related to the presented materials in the videos.

4.5 Measures of Learning Outcomes

The knowledge tests were reviewed by the presenter of the educational videos. The question types and difficulty levels were consistent with learning goals set for the subject. The test results were scored by the presenter. The knowledge tests contained two sections: retention (Mayer 2002), which perform the recall ability of students and transfer (Mayer 2002) for evaluating the problem-solving ability of the students after learning from the videos. All the questions are fill-in-the-blank questions (e.g., 'Four principles of creative fabric design_____ _____ and_____ (3 examples).' and filling in one blank correctly earns one point. The correct answers refer to the teaching content in the videos that '…So what principles does creative fabric design have? The first is understanding the

attributes of the fabric. Secondly, the fabric is fitted the design needs…'). There are 20 points for retention and 4 points for transfer. A student with all correct answers would score 24 points. All the questions closely relate to the video content, highlighted cues by the text communication through a pedagogical agent and a chat box in A-2, B-1 and B-2.

4.6 Measures of Eye Movement Data

We defined three areas of interest (AOI): the educational content displayed by PowerPoint presentations (AOI1), the chat box (AOI2) and the pedagogical agent (AOI3). In A-1, one AOI is identified with the educational content presented by PowerPoint. In B-1, two AOI are the educational content presented by PowerPoint and the pedagogical agent. In A-2, two AOI are the educational content presented by PowerPoint and the chat box. In B-2, three AOI are the educational content presented by PowerPoint, the pedagogical agent and the chat box (Figure 3).

Three types of eye movement data were collected: total AOI fixation duration time, AOI 1 total fixation duration time and total AOI fixation count. Total fixation duration time is the total amount of time of all fixation duration in a group of AOI. AOI 1 total fixation duration time is the total amount of time of the fixation duration in AOI 1. Total AOI fixation count is the amount of time that every participant took to look or fixate on the AOI for the first time, and it represents the visual searching speed for AOI (Rahal et al. 2019).

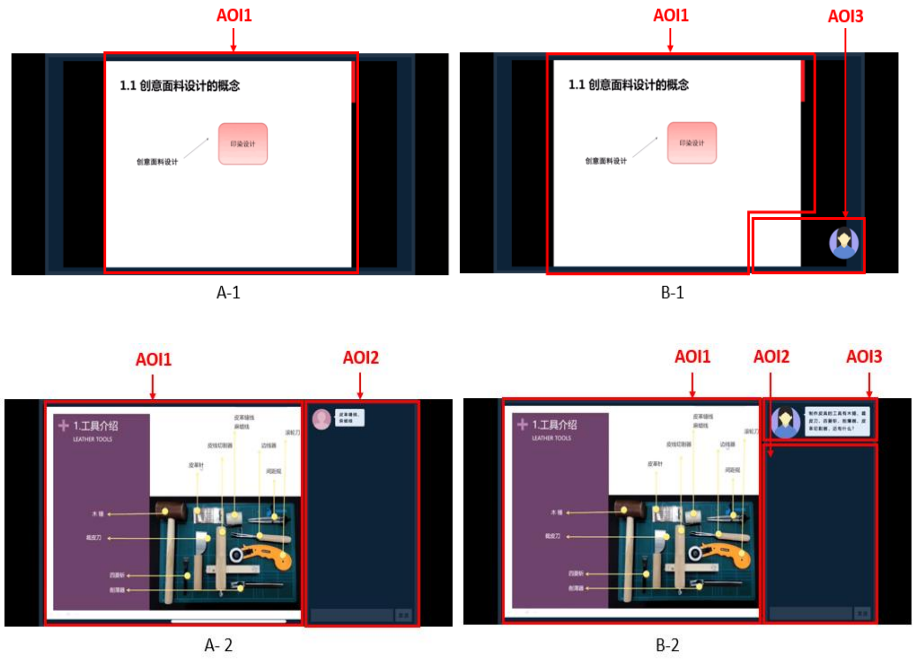

Figure 3. Areas of Interests (AOIs) in Four Conditions.

4.7 Procedure

Participants were required to complete a demographic questionnaire and prior knowledge test after the researchers explained the experiment content. Then participants were guided to sit in front of a

desk and learned from the presented videos shown on a laptop. During the learning process, the participants were told not to take notes and stay as still as possible. Each participant watched the two videos with different learning content. After each video, participants completed the knowledge test. After participants finished two videos and knowledge tests, a learning experience questionnaire was completed. The whole process lasted for half an hour.

5 Results

5.1 Retention and Transfer Tests

Across the four conditions, we compared the measures of learning performance including retention test and transfer test, the result of an age-based MANOVA test shows statistical significance (V=.07, F (1,103) =3.83, p=.025(<.05), η2 p=.07) for retention and transfer scores, another statistical significance is found in the results of media (four conditions; V=.22, F (1, 313) =4.25, p=.000(<.05), η2 p=.11). There is no statistical significance in the result of prior knowledge (V=.001, F (1, 103) =.041, p=.96, η2 p=.001). According to the results, generally, the treatments affect across four conditions, also the factor of age is assumed to affect the experiment. Tests of between-subject effects showed statistically significant differences in the retention test (F=.5.62, p=.001, η2 p =.139) but not in the transfer test (F=.0.58, p=.628, η2 p =.017) within four conditions. The results support that the transfer test is possibly insensitive to the medium, yet, retention tests are affected by the treatments (Table 2).

Table 2 Mean Scores and Standard Deviations on Learning Performance and Eye-Tracking Measures for Four Conditions

	A-1		B-1		A-2		B-2				
Dependent Variable	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**F**	**p**	η_p^2
Prior Knowledge Test	10	4	9	4	7	2	7	3	0.04	.96	.001
Retention Test	7.22	4.46	8.19	3.66	10.14	3.41	11.21	4.32	5.62	.001	.139
Transfer Test	3.15	1.03	3.04	1.09	2.79	1.23	2.89	.994	0.58	.628	.017
Learning Experience	14	5	15	5	14	5	14	4	0.15	.93	.004
Total Fixation Duration (s)	145.21	16.62	128.5	31.98	164.21	34.69	141.23	46.34	5.52	.001	.137
AOI 1 Total Fixation Duration (s)	145.21	16.62	110.11	26.5	148.69	33.8	114.29	40.1	12.7	.000	.268
Total AOIs Fixation Count	570.85	95.73	531.52	107	629.18	126.57	586.07	151.65	2.18	.095	.059

Note: A-1=control condition; B-2= Single-region and Pedagogical Agent; A-2= Multi-region Interface and Chatbox; B-2= Multi-region Interface, Pedagogical Agent and Chatbox; η_p^2*=partial eta-square.*

After conducting t-tests on retention for each of two conditions, the results reveal statistically significant differences between A-1 and A-2 (t(53)=-2.74, p=.008(<.05)), A-1 and B-2 (t(53)=-3.38, p=.001(<.05)), B-1 and A-2 (t(53)=-2.05, p=.045(<.05)) and B-1 and B-2 (t(53)=-2.8, p=.007(<.05) as shown in Table 3. The results show that retention scores are gradually increasing with treatments (Figure 4). However, since we ran the t-tests on retention six times, the data validity is reduced. For each significant p-value, we divided it by six and the refined p-value converted to .008, and the results remain statistically significant differences between A-1 and A-2 (t (53) =-2.74, p=.008(<=.008)), A-1 and B-2 (t (53) =-3.38, p=.001(<.008)) and B-1 and B-2 (t (53) =-2.8, p=.007(<.008). In the light of these t-test results, we assumed that the treatments including virtual pedagogical agents, and the chat box has a positive effect on retention scores.

5.2 Total AOIs Fixation Duration and Fixation Count

The measures of eye movements including total fixation duration (s) and total fixation count on the total area of interests and AOI 1 total fixation duration (s) are compared in four conditions. The age-based MANOVA test shows no statistical significance (V=.039, F (1,103) =1.39, p=.251, η2 p =.039) on total AOI fixation duration. AOI 1 total fixation duration and fixation count also show no statistically significant result in prior knowledge (V=.014, F (1, 103) =.474, p=.701 (=<.05), η2 p =.014). A statistical significance was found in the results of media (four conditions; V=.588, F (1, 313) =8.45, p=.000(<.05), η2 p=.196). According to the results, generally, the treatments affect across the four conditions, and the students' prior knowledge and age did not show coeffects in the experiment. Tests of between-subject effects present in-depth results of AOI fixation duration. AOI 1 total fixation duration and fixation count showed statistically significant differences in both AOI fixation duration and AOI 1 total fixation duration. The results suggest that AOI fixation duration and AOI 1 total fixation duration are affected by the treatments (Table 2).

Table 3 T-tests on Retention Test, Total Fixation Duration (s) and AOI 1 Total Fixation Duration(s) for Each Two Conditions

	Dependent Variable	F	t	df	Sig
A-1	Retention Test	1.1	-.868	52	.39
B-1	Total Fixation Duration (s)	7.28	2.4		.02
	AOI 1 Total Fixation Duration (s)	3.71	5.83		.000
A-1	Retention Test	2.89	-2.74	53	.008
A-2	Total Fixation Duration (s)	4.03	-2.58		.013
	AOI 1 Total Fixation Duration (s)	6.17	-.482		.632
A-1	Retention Test	0.2	-3.38	53	.001
B-2	Total Fixation Duration (s)	15.82	0.42		.676
	AOI 1 Total Fixation Duration (s)	12.54	3.71		.000
B-1	Retention Test	0.57	-2.05	53	.045
A-2	Total Fixation Duration (s)	0.03	-3.97		.000
	AOI 1 Total Fixation Duration (s)	.748	-4.7		.000
B-1	Retention Test	0.26	-2.8	53	.007
B-2	Total Fixation Duration (s)	3.03	-1.18		.243
	AOI 1 Total Fixation Duration (s)	3.71	-0.45		.651
A-2	Retention Test	1.32	-1.03	54	.307
B-2	Total Fixation Duration (s)	3.06	2.1		.04
	AOI 1 Total Fixation Duration (s)	1.01	3.47		.001

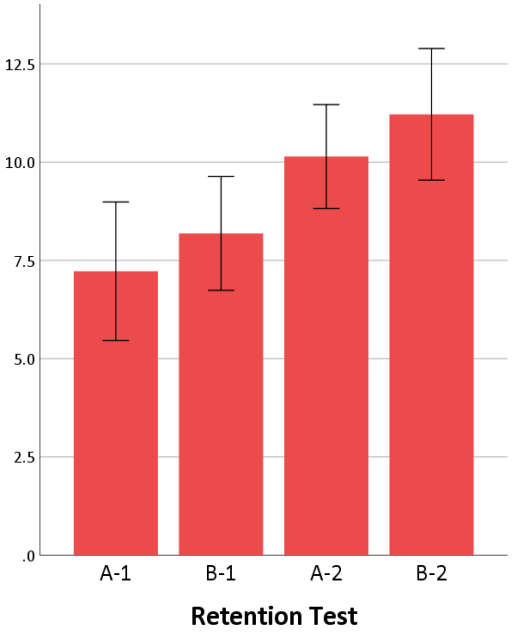

Figure 4. Differences in Retention Test across Four Conditions

T-tests were run on total AOI fixation duration between each pair of groups. The results reveal statistically significant differences between A-1 and B-1 (t(52)=2.4, p=.02(<.05)), A-1 and A-2 (t(53)=-2.58, p=.013(<.05)), B-1 and A-2 (t(53)=-3.97, p=.000(<.05)) and A-2 and B-2 (t(54)=2.1, p=.04(<.05); Table 3). T-tests were conducted on AOI 1 total fixation duration for each two conditions, statistically significant differences have been found in the comparisons of A-1 and B-1 (t(52)=5.83, p=.000(<.05)), A-1 and B-2 (t(53)=3.71, p=.000(<.05)), B-1 and A-2 (t(53)=-4.7, p=.000(<.05)) and A-2 and B-2 (t(54)=3.47, p=.001(<.05) as shown in Table 3. As the results show, total AOI fixation duration and AOI 1 total fixation duration fluctuate greatly with treatments as shown in Figure 5 and Figure 6. Since we ran the t-tests six times, the data validity was reduced. For each significant p-value, we divided it by six and converted the refined p-value to .008, and the results remain statistically

significant in total AOI fixation duration between B-1 and A-2 (t(53)=-3.97, p=.000(<.008)), and in AOI 1 total fixation duration between A-1 and B-1 (t(52)=5.83, p=.000(<.008)), A-1 and B-2 (t(53)=3.71, p=.000(<.008)), B-1 and A-2 (t(53)=-4.7, p=.000(<.008)) and A-2 and B-2 (t(54)=3.47, p=.001(<.008). According to the t-test results, we assumed that the treatments including virtual pedagogical agents and chat box effect on total AOI fixation duration and AOI 1 total fixation duration.

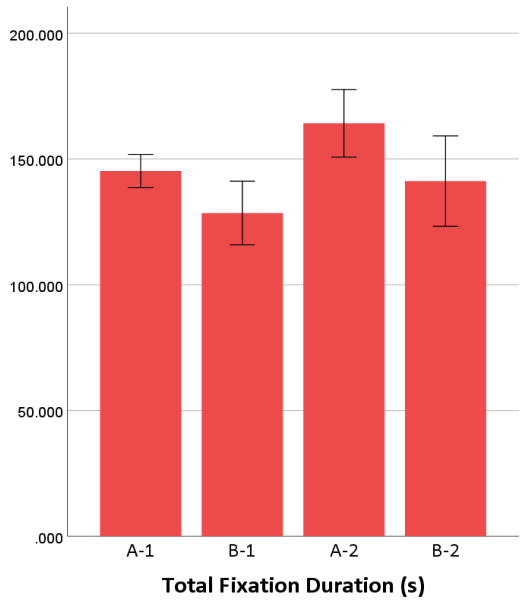

Figure 5. Differences in Total Fixation Duration (s) across Four Conditions.

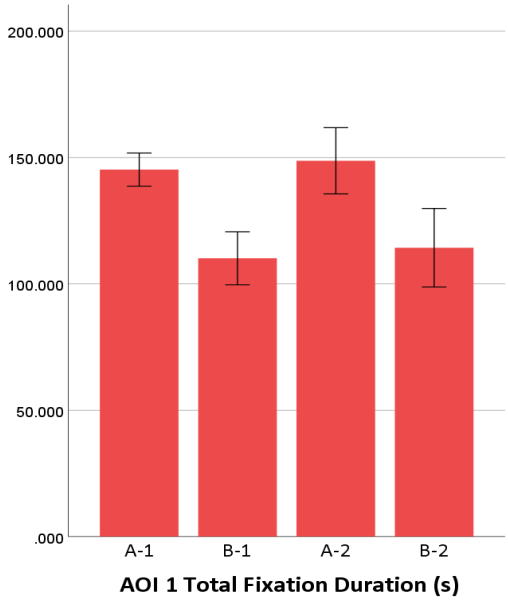

Figure 6. Differences in AOI 1 Total Fixation Duration (s) across Four Conditions.

6 Discussion and Future Direction

This study aims to explain the impact of multiple region interface layout, specifically pedagogical agents and the chat box effects on students' learning performance and visual attendance. The result based on the students' retention scores demonstrates that the students obtained the highest scores when the interface contained multiple regions, including presented materials, a pedagogical agent and a chat box. Furthermore, students earned the lowest retention scores when the interface was a single region with only presented materials. Meanwhile, there was no statistical significance of transfer scores within four conditions. The results suggest that students learn effectively in VLEs with multiple regions and two or more visual elements containing presented materials, pedagogical agents or a chat box. Moreover, the result shows that AOI 1 fixation duration significantly decreased when students learned from the VLEs with pedagogical agents. Students who learned from VLEs with pedagogical agents that provide text cues spent significantly less visual attendance on the presented materials but obtained higher retention scores than the students that learned from VLEs without pedagogical agents. We found that when students learned from a VLE that includes static pedagogical agents with text cues, the length of their eye fixation duration competes against the quality of learning performance. This result indicates that students spend quality time on learning

process and perform better retention scores under the condition of learning from the VLEs containing pedagogical agents. The VLEs designed with pedagogical agents are assumed to improve students' learning efficiency. Students that learned from the VLEs with a chat box received higher scores than the students learned from the VLEs without a chat box, but no significant difference in fixation duration. Chat boxes are designed to give scattered information but do not negatively affect learning performance. Thus, the results also indicate that both social interaction and multimedia in VLEs promote students' learning performance and the latter has more significant impact.

In the light of previous literature, visual attendance reveals a positive relationship with learning performance (Wang, et al. 2020; Stull, et al. 2018; Wang, et al. 2018); however, the correlation between visual attendance and learning performance are barely shown in this study. Corresponding to research questions, the findings of this study show the interventions of multiple regions layout, pedagogical agents and a chat box in VLEs positively influence learning performance, especially in retention knowledge. Aligned with previous literature, the result consistently examined that the VLEs with multimedia including pedagogical agents and a chat box improve students' learning performance. Additionally, it partially supports the social presence theory (Witmer & Singer 1998) that students show better learning performance in a social environment. However, the interpretation of data is comparably weak due to no significant result revealed in the students' learning experience questionnaire. The duration of visual attendance decreases when students learn from VLEs with the static pedagogical agents, which might be due to separated region layout (Mayer 2002), and conversely supports that students spend more quality time learning in the VLEs with multiple regions and static pedagogical agents. The correlation among students' preference, learning performance and visual attendance has the potential to be identified in a future study.

The limitations of this study are revealed in two aspects: one is that pedagogical agents provided educational information in text form, which, though intended to lead the students in conversational situations, initially generated redundant information. From the perspective of CTML (Mayer 2002), redundant information potentially influences the cognition learning process. Another limitation is the two types of teaching content (leather and textile design) potentially influence the results of knowledge tests. According to CTML (Mayer 2002), the design of multimedia learning is assumed to be spatial and have temporal contiguity. According to spatial contiguity effect, temporal contiguity effect and coherence effect (Mayer 2002), animated multimedia learning is the ideal design method for facilitating educational information dissemination and promoting knowledge applicability. For future studies, the design of pedagogical agents should be animated, constructing the representation of VLEs that are spatially and temporally integrated.

References

Anon, 2018. Proceedings of the 10th International Conference on Education Technology and Computers, ACM.
Bandura, A., 1977. *Social learning theory*, Englewood Cliffs, N.J.: Prentice Hall.
Beege, Maik et al., 2020. *Investigating the effects of beat and deictic gestures of a lecturer in educational videos*. Computers and education, 156, p.103955.
Biocca, Frank, Harms, Chad & Burgoon, Judee K, 2003. *Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria*. Presence (Cambridge, Mass.), 12(5), pp.456–480.

Castro-Alonso, Juan C et al., 2021. *Effectiveness of Multimedia Pedagogical Agents Predicted by Diverse Theories: a Meta-Analysis*. Educational psychology review, pp.Educational psychology review, 2021–01-03.
Ciolacu, Monica & Beer, Rick, 2016. *Adaptive user interface for higher education based on web technology*. 2016 IEEE 22nd International Symposium for Design and Technology in Electronic Packaging (SIITME), pp.300–303.
Chen, G., Zhou, S. and Zhi, T., 2019. *Viewing mechanism of lonely audience: Evidence from an eye movement experiment on barrage video*. Computers in Human Behavior, 101, pp.327-333.
Davis, Robert O & Vincent, Joseph, 2019. *Sometimes more is better: Agent gestures, procedural knowledge and the foreign language learner*. British journal of educational technology, 50(6), pp.3252–3263.
Dincer, Serkan & Doganay, Ahmet, 2017. *The effects of multiple-pedagogical agents on learners' academic success, motivation, and cognitive load*. Computers and education, 111, pp.74–100.
Fakhrunisa F, and Prabawanto S., 2020. *Online Learning in COVID-19 Pandemic: An Investigation of Mathematics Teachers' Perception*. In 2020 The 4th International Conference on Education and E-Learning (ICEEL 2020). Association for Computing Machinery, New York, NY, USA, 207–213.
Farhan, Wejdan et al., 2019. *E-learning systems versus instructional communication tools: Developing and testing a new e-learning user interface from the perspectives of teachers and students*. Technology in society, 59, p.101192.
Gillis, A. & Krull, L.M., 2020. COVID-19 Remote Learning Transition in Spring 2020: Class Structures, Student Perceptions, and Inequality in College Courses. Teaching sociology, 48(4), pp.283–299.
Gnewuch, U., Morana, S. and Maedche, A., 2017, December. *Towards Designing Cooperative and Social Conversational Agents for Customer Service*. In ICIS.
Goldin-Meadow, Susan & Beilock, Sian L, 2010. *Action's Influence on Thought: The Case of Gesture*. Perspectives on psychological science, 5(6), pp.664–674.
Horovitz, Tal & Mayer, Richard E, 2021. *Learning with human and virtual instructors who display happy or bored emotions in video lectures*. Computers in human behavior, 119, p.106724.
Hu, Tsu-Wu et al., 2019. *Research on Interface Improvement of English Vocabulary Learning APP*. Proceedings of the 2019 7th International Conference on information technology: iot and smart city, pp.165–169.
Hung, I.C. and Chen, N.S., 2018. *Embodied interactive video lectures for improving learning comprehension and retention*. *Computers & Education*, *117*, pp.116-131.
Koning, Björn B, Rop, Gertjan & Paas, Fred, 2020. *Effects of spatial distance on the effectiveness of mental and physical integration strategies in learning from split-attention examples*. Computers in human behavior, 110, p.106379.
Lam, Winnie et al., 2019. *Investigating Online Collaborative Learning on Students' Learning Outcomes in Higher Education*. Proceedings of the 2019 3rd International Conference on education and e-learning, pp.13–19.
Lawson, Alyssa P et al., 2021. *Do Learners Recognize and Relate to the Emotions Displayed By Virtual Instructors?* International journal of artificial intelligence in education, 31(1), pp.134–153.
Li, Wenjing et al., 2019. *Getting the Point: Which Kinds of Gestures by Pedagogical Agents Improve Multimedia Learning?* Journal of educational psychology, 111(8), pp.1382–1395.
Makransky, Guido et al., 2019. *Equivalence of using a desktop virtual reality science simulation at home and in class.* PloS one, 14(4), p.e0214944.
Makransky, Guido, Wismer, Philip & Mayer, Richard E, 2019. *A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation*. Journal of computer assisted learning, 35(3), pp.349–358.
Mayer, R.E., 2002. *Multimedia learning*, New York: Cambridge University Press.
Mirsarraf, M.R., Khazaei, M. and Ahmadpanah, A., 2017. *Role of pictorial symbols in the web interface design*. Int J Arts, 7, pp.17-22.
Moreno, Roxana & Mayer, Richard, 2007. *Interactive Multimodal Learning Environments: Special Issue on Interactive Learning Environments: Contemporary Issues and Trends*. Educational psychology review, 19(3), pp.309–326.
Moreno, Roxana & Mayer, Richard E, 2004. *Personalized Messages That Promote Science Learning in Virtual Environments*. Journal of educational psychology, 96(1), pp.165–173.
Rahal, R.M. and Fiedler, S., 2019. *Understanding cognitive and affective mechanisms in social psychology through eye-tracking*. Journal of Experimental Social Psychology, 85, p.103842.

Rapanta, Chrysi et al., 2020. *Online University Teaching During and After the Covid-19 Crisis: Refocusing Teacher Presence and Learning Activity*. Postdigital Science and Education, pp.Postdigital Science and Education, 2020–07-07.
Rey, Günter Daniel, 2014. *Seductive details and attention distraction – An eye tracker experiment*. Computers in human behavior, 32, pp.133–144.
Russ, S and Hamidi F., 2021. *Online learning accessibility during the COVID-19 pandemic*. In Proceedings of the 18th International Web for All Conference (W4A '21). Association for Computing Machinery, New York, NY, USA, Article 8, 1–7.
Schroeder, Noah L, Adesope, Olusola O & Gilbert, Rachel Barouch, 2013. *How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review*. Journal of educational computing research, 49(1), pp.1–39.
Shin K, and Downing K., 2011. *User centred e-learning platform design*. In Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces (DPPI '11). Association for Computing Machinery, New York, NY, USA, Article 8, 1–4.
Short, J., Williams, E. & Christie, B., 1976. *The social psychology of telecommunications*, London: Wiley.
Stull, Andrew T, Fiorella, Logan & Mayer, Richard E, 2018. *An eye-tracking analysis of instructor presence in video lectures*. Computers in human behavior, 88, pp.263–272.
Sutton, Rosemary E & Wheatley, Karl F, 2003. *Teachers' Emotions and Teaching: A Review of the Literature and Directions for Future Research*. Educational psychology review, 15(4), pp.327–358.
Swan, K., 2004. Issues of interface. *European Journal of Open, Distance and E-learning*, *7*(1).
Swan, K., 2004. *Learning online: A review of current research on issues of interface, teaching presence and learner characteristics*. Elements of quality online education: Into the mainstream, 5, pp.63-79.
Sweller, J., 1994. *Cognitive load theory, learning difficulty, and instructional design*. Learning and instruction, 4(4), pp.295-312.
Turumogon, P. and Baharum, A., 2018. *Identifying a user interface web design standard for higher learning institutions using Kansei engineering*. Indonesian J. of Electrical Engineering and Computer Science, 11(1), pp.90-97.
Tu, Chih-Hsiung, 2000. *On-line learning migration: from social learning theory to social presence theory in a CMC environment*. Journal of network and computer applications, 23(1), pp.27–37.
Wang, Fuxing et al., 2018. *Animated Pedagogical Agents as Aids in Multimedia Learning: Effects on Eye-Fixations During Learning and Learning Outcomes*. Journal of educational psychology, 110(2), pp.250–268.
Wang, Jiahui & Antonenko, Pavlo D, 2017. *Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning*. Computers in human behavior, 71, pp.79–89.
Wang, Xue et al., 2020. *Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos*. Computers in human behavior, 107, p.106279.
Wang Y. Zhu X. and Yang Z. 2018. *Design and development of online practice teaching platform of new information technologies*. In Proceedings of the 10th International Conference on Education Technology and Computers (ICETC '18). Association for Computing Machinery, New York, NY, USA, 321–325.
Wang Z., and Cruz F. I., 2020. *Analysis of the Impact of COVID-19 on Education Based on Geotagged Twitter*. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19 (COVID-19). Association for Computing Machinery, New York, NY, USA, 15–23.
Witmer, Bob G & Singer, Michael J, 1998. *Measuring Presence in Virtual Environments: A Presence Questionnaire*. Presence (Cambridge, Mass.), 7(3), pp.225–240.
Xie, Heping et al., 2019. *Coordinating Visual and Auditory Cueing in Multimedia Learning*. Journal of educational psychology, 111(2), pp.235–255.
Vlasenko, K., Kovalenko, D., Chumak, O., Lovianova, I. and Volkov, S., 2020. *Minimalism in Designing User Interface of the Online Platform" Higher School Mathematics Teacher"*. In ICTERI Workshops (pp. 1044-1057).
Volery, Thierry & Lord, Deborah, 2000. *Critical success factors in online education*. International journal of educational management, 14(5), pp.216–223.
Zeng L., 2020. *Which Kind of Learning Form Do Students Want in China during COVID-19 Outbreak*. In 2020 The 4th International Conference on Education and Multimedia Technology (ICEMT 2020). Association for Computing Machinery, New York, NY, USA, 51–56.