





Creating Dialogue Between a Tutee Agent and a Tutor in a Lecture Video Improves Students' Attention

Ari Nugraha , Izhar Almizan Wahono, Jianpeng Zhanghe, Tomoyuki Harada, and Tomoo Inoue 

University of Tsukuba, Tsukuba, Japan

{ari.nugraha, inoue}@slis.tsukuba.ac.jp, {s1826099, s1921659, s1721687}@s.tsukuba.ac.jp

Abstract. One of the challenges found with learning using a lecture video is the short attention span of students due to high cognitive load when they are watching the video. Several methods have been proposed to increase students' attention, such as segmenting learning content into smaller pieces where each piece has short video duration. In this study we proposed an enhanced monologue lecture video with a tutee agent to mimic the dialogue between a tutor and a tutee to increase student attention to the lecture video. Based on lab evaluation including eye fixations data from an eye tracker, the videos enhanced with the tutee agent make the students' attention more frequent to the learning material presented in the lecture video and at the same time lowering their attention span to it.

Keywords: Tutee · Tutee agent · Pedagogical agent · Dialogue video · Lecture video · Student attention

1 Introduction

Delivering lecture in a form of video format becoming quite common nowadays as tools for creating a video already accessible where one can even produce just using their smartphone. While the Web video technology and standard getting more mature, in last decade we also saw the rise of massive open online courses (MOOC) and open courseware where its main way for learning content delivery is by the means of lecture video. MOOCs and open coursewares provided by several world' high ranking university such as the MIT, opening opportunity for everyone at their own time and their own pace. In this sense, MOOC and open courseware provide greater accessibility and time convenience for learners. However, even though learning through MOOC provides flexible way of learning for the enrolled students, problems such as high cognitive load caused by information overload is still faced by learners. High video dropout rates caused by too long videos duration, and boredom due to short attention spans are another problem in watching lecture video delivered in distance learning platform. In this study, we proposed an enhanced lecture video material with embedded tutee agent to make the lecture video more exciting to watch by the students and in the end increasing student focus with the learning content in the video.

2 Literature Review

2.1 Video as Learning Material

The emergence of MOOCs and Open Coursewares, brings thousands of online courses being offered in the Internet. Each of these online courses consists of several parts where most of the content is recorded lecture video as a learning material. One of the reason why video is still becoming the *de facto* format for delivering learning content in distance learning environment is that a video provides a rich and powerful medium and can present information in an attractive and consistent manner [1]. Another benefit of lecture video is that it can be well suited to visualize the abstract or hard-to-visualize phenomena that are important in many science classes such as biology [2]. Outside of MOOC's realm, the usage of lecture video as a learning material in traditional classroom are becoming popular as instructors are increasingly making use of flipped classrooms method, whereby students are encouraged to watch the recorded lecture video on their own time and engage in activities geared toward a more in-depth understanding of the subject matter in the classroom [3]. While videos on MOOCs are structured inside courses, popular video sharing application such as YouTube also offers unstructured educational video provided not only from professional tutor, but also from non-professional content creators.

One of important thing to be considered when delivering the learning material in a video format is how effective the video in maximizing student learning. In the literature review on principals of effective lecture video, Brame [2] suggests three elements for video design and implementation that can help instructors maximize video's utility: 1) cognitive load, 2) student engagement, and 3) active learning. The cognitive theory of multimedia learning defines two channels for information acquisition and processing: a visual channel and an auditory channel [4]. The use of these two channels can facilitate the integration of new information into existing cognitive structures and lecture videos are the perfect medium for this, as it provides both visual and auditory information at the same time.

2.2 Learning with Dialogue Video

Other factor that can affects student engagement in lecture video according to the study is how the tutor or teacher presented the learning material or lecture style inside the video. Several lecturing style in the video exists such as the Khan-style or the talking-head style. The talking-head style with the tutor's head appears in the video has positive impact to the student enjoyment and learning performance compared to paper book style [5]. Lecture video style which involves a dialogue between a tutee and student also proved to have positive impact to the students learning. In this case the student become an observer student who learning by watching the dialogue. Study found that the students who watched the dialogue video have better constructive and interactive behaviors compared to watching monologue-videos and they are benefited from the presence of the tutees as they pay more attention to what the tutees said than to what the tutors said [6]. The advantage of this method is that it even does not need to be conducted with professional tutors; a meta-analysis study found that people with untrained tutoring skills and had moderate domain knowledge could become tutor [7].

2.3 Role of Pedagogical Agent in Learning

A large body of research related to the pedagogical agent in learning environment exists. Regarding the student's focus and attention, pedagogical agent, which usually represented as a human-like character, can provide instructional support and motivational elements into multimedia learning [8]. Pedagogical agent's voice and representation will produce social signals that trigger social responses from students. Study on influence of learner's social skill in collaboration with pedagogical agent showed that learners with higher social skills performed better on the explanation task with the agent than those with lower social skills and this is an indication that learners perceived their interaction with the agent as same with human-human interaction, where learners with low social communication skills have difficulty in collaborative learning [9].

As the most studies on pedagogical agent define the role of an agent as a tutor to the student, an agent role as a tutee also have been explored. Students who overheard a dialogue between a virtual tutee and a virtual tutor, learned more, took more turns in mixed-initiative dialogue, and asked more questions than those in the monologue-like condition where only a virtual tutor exists [10]. In more recent study, researchers have proposed a virtual tutee system (VTS) to improve college students' reading engagement, where the students take on the role of tutor and teach a virtual tutee. The study found that the students in VTS group engaged in a deep level of cognitive processing and have higher reading performance than students in the non-VTS group [11].

One of the benefits of using animated pedagogical agent in multimedia learning is that it is possible to more accurately model the dialogs and interactions that occur during novice learning and one-on-one tutoring [12]. Based on these previous studies of pedagogical agent and combined with the benefits of dialogue between a tutee and a tutor in the lecture video, in this study proposed a prototype tool to add a tutee agent as a replacement for human tutee in monologue video.

3 Proposal

3.1 Tool Overview

To add a tutee agent inside the lecture video with our prototype tool, a teacher or course designer needs to make an annotation on the designed lecture video with our tutee agent annotation interface. The annotation texts created by the teacher serve as the tutee agent utterances inside the video and will be uttered by the tutee agent at specified time in the video. The tutee agent voice is created by utilizing text-to-speech cloud service. The steps to do this as illustrated in Fig. 1 are:

1. Upload the designated lecture video
2. Add important metadata to the uploaded lecture video
3. Create annotation texts to the uploaded lecture video by selecting specific time in the video using the Tutee Agent's Annotation interface.

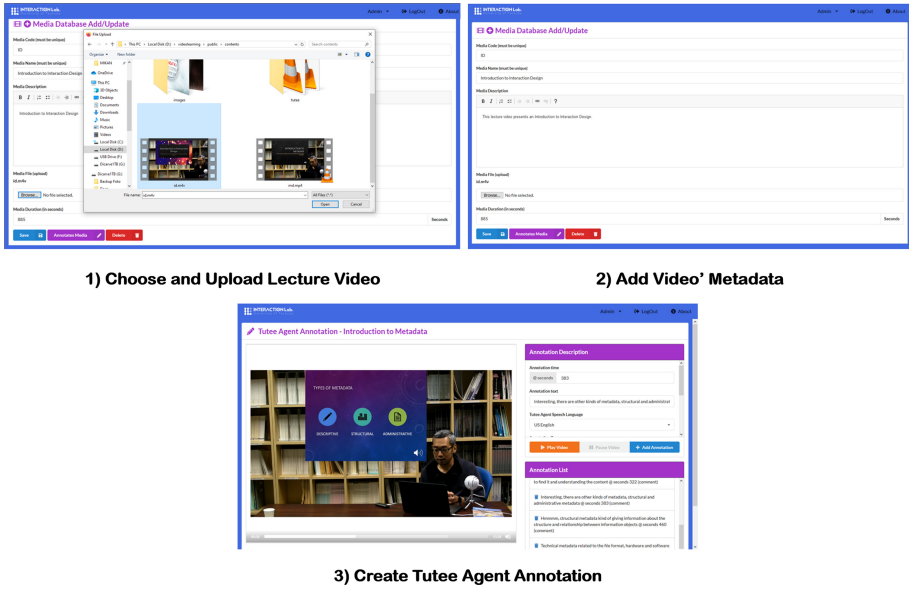


Fig. 1. Lecture video and Tutee Agent Annotation workflow

3.2 Lecture Video Produced by the Tool

The result of the annotated lecture video is a lecture video with the tutee agent embedded into the lecture video. For the animated tutee agent itself, we created in animated PNG format as it provides high resolution image and can be embedded easily in Web environment. Using Web compatible format tutee agent also make customization of tutee agent easier for future development and study. When the tutee agent making a specific utterance at certain time in the video, the animated tutee agent animation is synchronized with the related voice generated from the cloud service while the lecture video is paused to mimic turn taking conversation between tutor and tutee agent. A text balloon containing the annotation text also showed up above the tutee agent while she is speaking.

4 Evaluation

4.1 Purpose

To evaluate the effect of lecture videos enhanced by our proposed tool, we have conducted a user study using within-subject experiment design by involving participants as an observer student. In this study, participants watched lecture videos using video user interface of our system in two conditions: 1) Monologue style (MON) where participants watch a monologue lecture video, where only a human tutor present inside the video and, 2) Dialogue style with tutee agent (DIA) where participants watch a monologue lecture video with a tutee agent added in the video. As a counterbalance for learning effect and bias from the learning style, we order the experiment condition differently for each participant. Before the experiment conducted, we have been granted an ethical approval from our university' IRB.

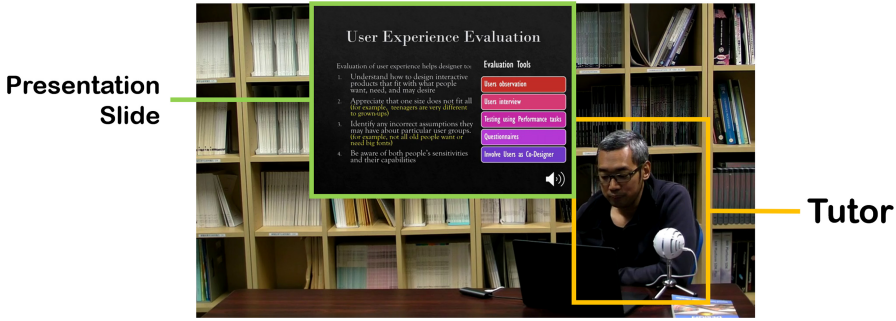


Fig. 2. The monologue lecture video as a learning material in this study delivered in talking head with presentation. In talking-head style, the tutor’ head is appeared inside a lecture video explaining learning content.

4.2 Experiment Apparatus and Settings

In this experiment we used a personal computer (PC) connected to our tool and the tool was accessed via web browser. Participants watched the lecture video using the PC with 24-in. Full HD (1920 × 1080 resolution) monitor in our laboratory and listened to the learning material using headset. To collect data on user attention and eye fixation, we equipped participants with the open source Pupil eye-tracker [13] (see Fig. 3). While watching the lecture video, participants were seated in front of the PC’s monitor and we asked participants to adjust their seat position as convenient as they want. Three video cameras in different angles are used in this experiment to record participant’s behavior while watching the lecture video.

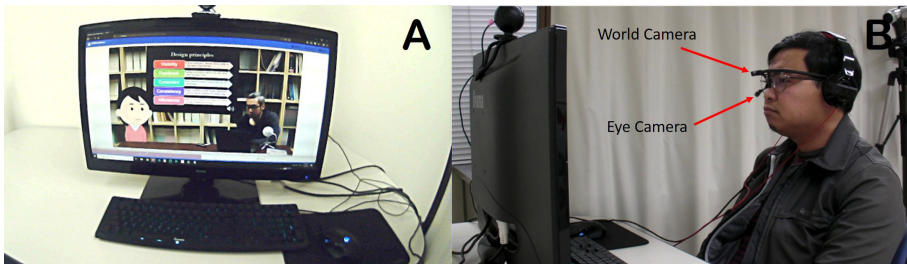


Fig. 3. A) Participant’s world view from the Pupil’s world camera represent what participant saw with their eye, B) Participant in our experiment seated in front of PC’s monitor wearing the Pupil eye tracker and headset while watching the lecture videos.

Experiments with eye-tracker commonly restrict the head movement of participants using a chin rest to reduce head movements and ensure a constant viewing distance [14], we do not using this kind of tool in our experiment as it is not reflecting an ideal condition when student watching lecture video. We did not limit the movement of participants’ head and body when they were watching the lecture video. The purpose for this was that

we wanted our participants to be as convenient as possible while watching the lecture video and to reflect real world condition in watching lecture video.

4.3 Materials

4.3.1 Lecture Videos

For the learning materials used in our evaluation, we produced two lecture videos with different themes in a monologue style. These monologue style video then enhanced into dialogue style for the DIA condition, where the tutee agent presents inside the video and make questions and comments related to what the tutor explains in the video. The tutee agent utterances inside the video were scripted and annotated using our tutee agent annotation interface. The themes of the lecture videos produced are: 1) Introduction to Interaction Design (ID), and 2) Introduction to Metadata (MD). The learning content for ID theme video is taken from *The Interaction Design* textbook, while for the MD theme video is taken from the modified and simplified version of Metadata MOOC series videos from YouTube by Professor Jeffrey Pomerantz of School of Information and Library Science, The University of North Carolina of Chapel Hill (<https://www.youtube.com/watch?v=fEGEJhJzrB0>).

To investigate how the tutee agent could affect in variety type of video content, we designed The ID theme video presentation slides to contain more text than the MD theme video as it is known from limited capacity theory states that our short-term memory is quite small; therefore, offering too much information on the slides such as too many text could lead to a high cognitive load and split attention [15]. Both of these lecture materials was delivered using talking-head with presentation style (Fig. 2) as it is one of the popular learning content delivery style in lecture video [5]. To strike the balance between video duration and learning content, we designed both lecture videos to have duration up to 15 (fifteen) min.

4.3.2 Tutee Agent Utterances

Studies on pedagogical agent have shown that by employing strategic utterances in dialog, such as asking for an explanation, repeating learning content, or providing suggestions, could stimulate reflective thinking and the metacognition processes involved in understanding [9]. Based on this notion of strategic utterances of pedagogical agent, we annotated the tutee agent at specific time of video. As slide transition in presentation can be indicated as a transition to different learning topic, we were making our tutee agent to make utterances at the specific time on the lecture video where it is near the end of every slide or before the slide transition occurred.

By making the tutee agent to utters summary to the current slide content and tutor's explanation, we implemented the concept of repetition. Repetition can be one of the virtual tutee contribution in the dialog by uttering key concepts mentioned by the tutor in his previous conversational turn [16]. Example of this repetition can be seen on Table 1.

Table 1. Example of the tutee agent summarization in MD theme video

Tutor	<i>“... from library and librarian we learn how to describe something effectively. Effectively here means for the purpose of storage and later retrieval by the user of library. Information about the content of a thing or resource that aids in finding or understanding it, is referred to as descriptive metadata. If you are looking for a book, the most obvious thing you care as a reader are, first, what the book called or the title, second, who wrote it or the author and third what is the book about or the subject! Descriptive metadata describes the book using those three aspects”</i>
Tutee agent	<i>“I see, descriptive metadata describes the object that can help people to find it and understanding the content”</i>

Besides making repetition of the tutor explanation in the form of conclusion, we also designated our tutee agent to utter questions. The question from the tutee then answered by the tutor in the next turn of explanation. An example of question utterance from the tutee agent is: *“...Anything? what kind of things can be described?”*. This tutee’ question is placed before the tutor changing his explanation about type of thing that can be described in metadata. Right after the tutee’ utterance ended, the video playing again, and the tutor start to continue his explanation as can be seen on Table 2. This kind of interaction of the tutee agent is also form of virtual tutee contribution to the learning by making the observer students anticipated the next learning content explanation by the tutor [16].

Table 2. Example of the tutee agent utters question to the tutor

Tutee Agent	<i>“...Anything? what kind of things can be described?”</i>
Tutor	<i>“The thing being described maybe natural, or an artificial object, physical or digital object, whatever It also can be an information object, such as webpage, a book, an article, a historical manuscript, etc.”</i>

4.4 Participants

In this experiment, we recruited twelve (N = 12) participants who were all undergraduate and graduate students. As our participants watching the dialog between a tutee agent and a tutor, they are becoming a dialog observer or an observer student.

4.5 Procedure

At the beginning of the experiment, we gave an explanation to the participants about the experiment and the consent paper. After the participant agreed with the experiment condition they read and filled the experiment consent form. Each participant experienced the two experiment conditions, DIA and MON, or MON and DIA, depends on the

participant task order document we have made before. We equipped participants with the Pupil eye-tracker to collect data on their eye-fixation points. Our Pupil eye-tracker device is embedded with two cameras, first is the world camera which represent what participant saw with their eye and the second one is the eye-camera which detect participant's the eye movement.

Before participants start watching the lecture videos, we carried out the Pupil eye-tracker calibration procedure to make sure the eye-tracking process run accurately. Using the calibration feature provided by the Pupil' software suite, the Pupil eye-tracker calibration was done by asking a participant to follow five dots which are appeared on the screen without moving their head to make sure the Pupil' eye camera could detect the movement of participant' pupil correctly. We performed this procedure several times for each participant to make sure that the Pupil eye-tracker device has high confidence detection between 0.7 to 1.0 (maximum) value. Participants watched the lecture videos in one pass without re-watch and they were not allowed to control the pace of video by pause, seek backward, or forward the videos. After finished watching each video, the participant then filled a questionnaire related to their user experience and then answered free recall test related to the learning content in the last watched video.

5 Result

5.1 Analysis of the Data

Learning Gain

To measure learning gain particularly to short-term memory retention, we collected and calculated data from participants' free recall answer sheets. For each video theme, we created six deep questions related to learning content in the video. From these participant' answers sheet, we counted how many relevant, related, and irrelevant propositions. A relevant proposition can be a phrase or sentences which matched or has the same meaning with what tutor explained or tutee uttered in the video. A related proposition is a phrase or sentence which not too relevant to the question asked, but still have some relation with the right answer, and an irrelevant proposition is an answer which does not have any relevance with the question or incorrect answer.

Eye Fixation Frequency and Duration

To measure and analyze our participants attention to the lecture videos from the eye tracker video recordings and raw data, we followed four steps procedure for each video (Fig. 4).



Fig. 4. Workflow of eye fixation data collection and analysis

The procedure was as follow: 1) Collected eye fixation points raw data and videos recorded from the eye-tracker device using the Pupil Capture apps. 2) Each of these videos (in HD 1280 × 720 resolution) were then exported using Pupil Player apps into a rendered video which contain a yellow circle marker (generated automatically by the Pupil Player app based on the raw data). The yellow circle marker in the rendered video move based on the participant’s eye fixation movements (Fig. 5A). 3) The rendered video then added into ELAN [17] annotation tool in which the eye-fixation periods of participant were annotated. 4) We then conducted statistical tests for the eye fixation occurrences frequency and the duration of eye fixation based on these annotations result.

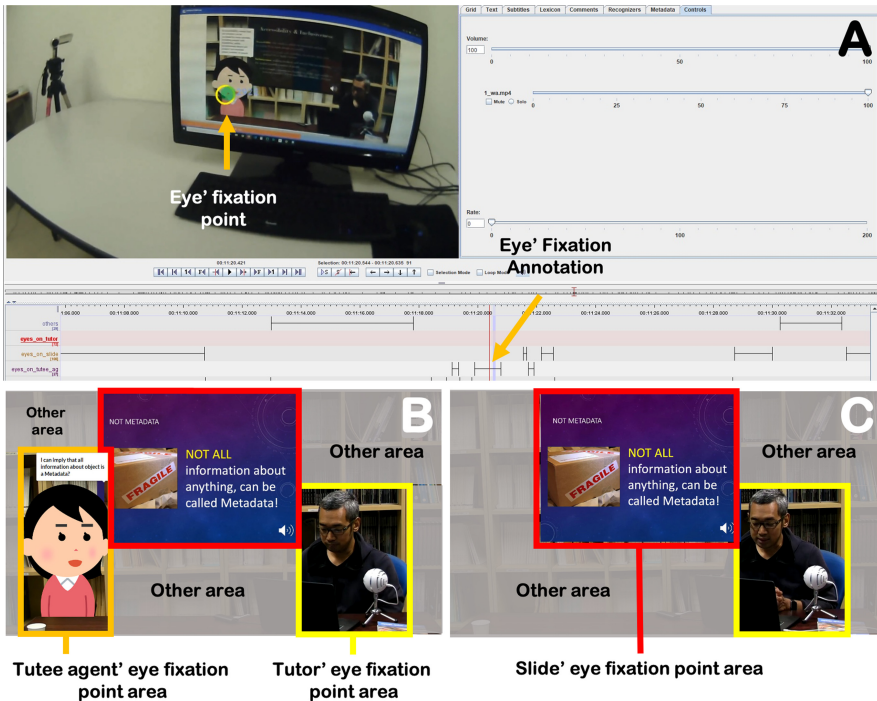


Fig. 5. A) Annotation of participant’ eye fixation point (indicated by yellow circle) in the video using ELAN. B) Annotation’s AOI where the eye fixation points landed on for the dialogue video. C) Annotation’s AOI for the monologue video. (Color figure online)

The frequency of eye fixation was measured on how many times the eye of participant switch fixation to the one of AOI. For example, we counted one eye fixation to the tutor area after it is moved from the slide area to the tutor area. In doing the eye fixation points' video annotations, we defined three main areas of interest (AOI) on the video where the eye fixation landed on. These AOI are the tutee agent area, tutor area and the slide area as can be seen on Fig. 5B. If participants eye fixation landed on another area than these AOI, we annotated them as "other". We then counted how many times (the frequency) the eye fixation landed inside these areas and how long (the duration) the fixation points last.

Interview

To gain deeper understanding on the effect of the tutee agent in dialogue video which otherwise not reflected in learning gain or attention from eye-tracker, we conducted semi-structured interview at the end of each experiment session. The interview was audio recorded by the experimenter and then transcribed. When then making our analysis based on the transcript data.

5.2 Learning Gain

Based on the number of relevant, related, and irrelevant propositions from participants' answers, we measured the difference between DIA (dialogue) and MON (monologue) condition (Fig. 6).

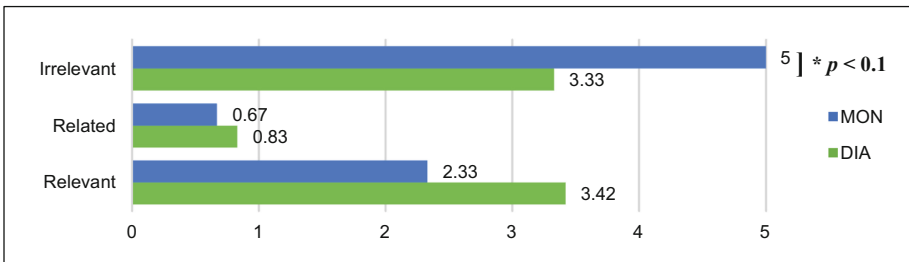


Fig. 6. The number of relevant, related, and irrelevant propositions

Our MANOVA test reported non-significant effect of video style/condition (DIA & MON) as the independent variable to all three independent variables (Relevant, Related & Irrelevant proposition) with $F(3, 20) = 1.40, p = .27$; Wilk's $\Lambda = .826$. Univariate test reported marginal effect of video style to the Irrelevant variable result where participants in the DIA condition were writing less irrelevant proposition compared to the MON condition, $F(1, 22) = 3.71, p = .06$.

5.3 Eye Fixation Frequency

The total eye fixation recording videos which we annotated were 20 videos from the total of 24 videos (each participant has two eye tracker video recordings). We excluded

4 videos from this analysis because of incomplete data from 2 participants. One video in the dialogue condition was not containing any eye-fixation marker because the eye-tracker failed to detect the participant eye' pupil in the middle of experiment, despite we already performed several calibration procedures. Another one in monologue condition resulted in incomplete eye fixation data because half of the eye tracker recording video was corrupted (Fig. 7).

Table 3. The average occurrence of eye fixations per minute

Fixation AOI	DIA (N = 10)		MON (N = 10)		MS	F	df	p
	M	SD	M	SD				
Slide	9.86	3.34	6.08	2.45	71.21	8.27	1	** .010
Tutor	8.73	4.23	5.89	3.27	40.38	2.81	1	.110
Other	2.36	1.74	3.47	1.36	6.11	2.48	1	.133
Agent	4.01	1.65	-	-		-	-	-

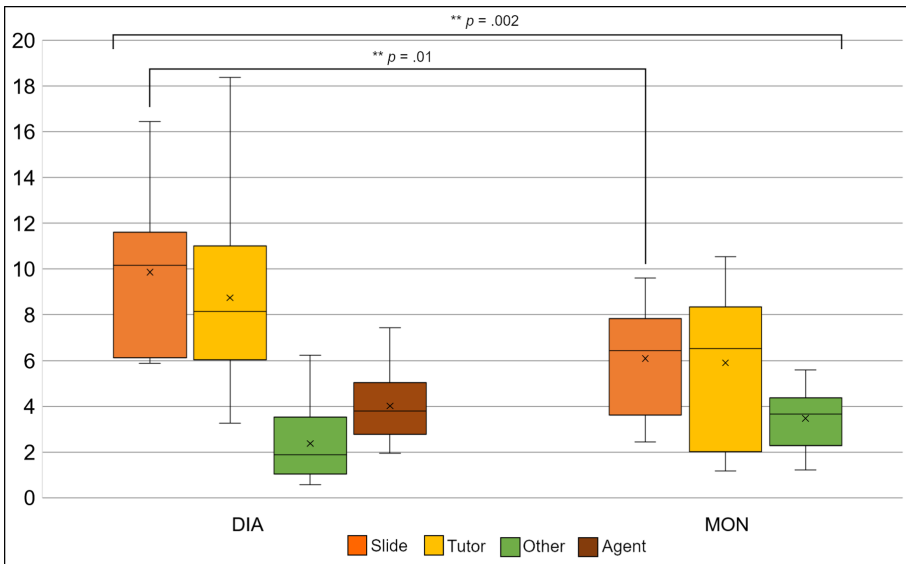


Fig. 7. Eye-fixation occurrence per minute for DIA and MON condition.

For the analysis to the eye fixation frequency, first we present the result on the eye fixation frequency to the three main AOI in both DIA and MON conditions. These numbers represent the number of eye fixation per minute (fpm) from the beginning until the end of the lecture video. We conducted ANOVA to compare the effect between video style (DIA and MON) for the three fixation AOI. The multivariate test reported

a significant effect of video style with $F(3, 16) = 8.03, p = .002$; Wilk's $\Lambda = .399$. Furthermore, univariate tests showed there was a significant effect of video style where participants have higher frequency of eye fixations to the slide area in the DIA style than the MON style, $F(1, 18) = 8.27, p = .010$. However, as can be seen on Table 3, no significant differences were found in the tutor area ($F(1, 18) = 2.81, p > .05$) and the other area ($F(1, 18) = 2.48, p > .05$).

5.4 Eye Fixation Duration

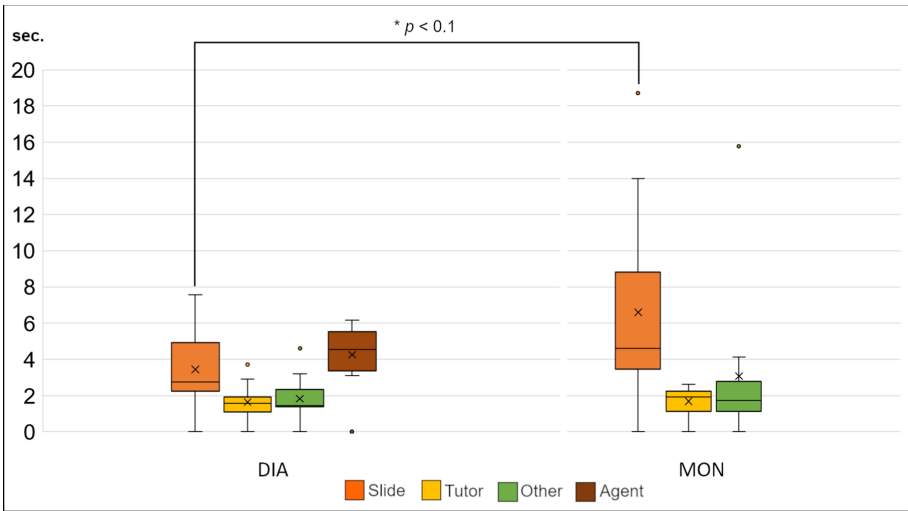


Fig. 8. Duration of each eye fixation to all AOI in DIA and MON condition.

Table 4. Duration of each eye fixation on AOIs (sec.)

Fixation AOI	DIA (N = 10)		MON (N = 10)		MS	F	df	p
	M	SD	M	SD				
Slide	3.78	1.86	7.25	5.17	60.06	3.96	1	*.062
Tutor	1.80	.88	1.86	.63	.01	.03	1	.861
Other	2.01	1.13	3.37	4.47	9.30	.87	1	.362
Agent	4.69	1.09	–	–	–	–	–	–

Beside from the number of times the participants fixated their eye to the three areas, we also try to look at the average duration of eye fixation between condition. Figure 8 illustrates the comparison of data spread between DIA and MON condition and all AOI. Our multivariate test reported non-significant effect of the video style to all fixation

AOIs. As can be seen on Table 4, univariate test reported marginal difference between video style where participants in MON have longer duration to the slide area, $F(1, 18) = 3.96, p = .062$. This result indicated that the attention of participants tended to more keep on the slide area in the monologue lecture video (Fig. 9).

5.5 Eye Fixation Rate

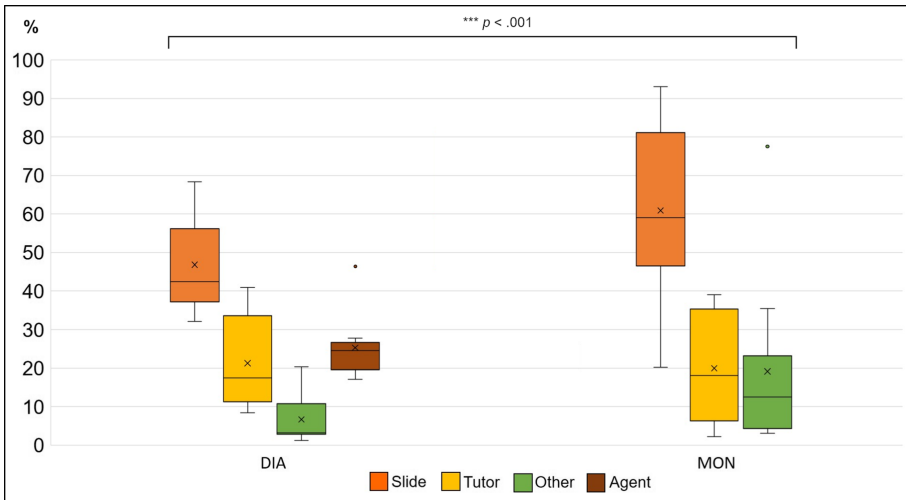


Fig. 9. Eye fixation rate (%)

Table 5. Eye fixation rate (%)

Fixation AOI	DIA (N = 10)			MON (N = 10)			MS	F	df	p
	M	SD	Rate (%)	M	SD	Rate (%)				
Slide	28.10	7.61	46.81	36.53	13.67	60.9	356.72	2.91	1	.105
Tutor	12.73	7.06	21.23	11.97	8.16	19.94	2.98	.05	1	.824
Other	4.01	3.61	6.69	11.50	13.61	19.16	280.05	2.82	1	.110
Agent	15.16	4.96	25.27	–	–	–	–	–	–	–
Total	60	–	100	60	–	100	–	–	–	–

To obtain the rate of eye fixation duration in each minute, we multiply the frequency of eye fixation per minute and the duration of each eye fixation. We normalized the data, so the total time is exactly 60 s for all AOI in each condition. Our MANOVA reported significant effect of video style, $F(3, 16) = 29.05, p < .001$; Wilk’s $\Lambda = .155$. However univariate tests on each area as detailed on Table 5, did not reported any effect of video style on all AOIs.

6 Discussion

As we have described on the result part before, result from learning gain showed no significant difference in learning gain, particularly to the number of relevant propositions. This result might be caused by the tutee agent did not provide enough scaffolding to the memory retention of participants in watching the lecture videos. Study from Mayer on the presence effect of pedagogical agent found that students who learned with an agent on the screen did not get better results compared to students who study without agent on the screen [18].

From eye fixation data collected from the eye tracker, our data analysis found when participants watched the dialogue video with the tutee agent, they were tended to more frequently paid attention to the slide area. This might be an indication that the tutee agent utterances make participants to review the presentation slide inside the video more often than in the monologue video. However, result from the duration for each eye fixation in the monologue video revealed that the participants spent their attention longer to the slide area. In the dialogue video, the tutee agent took some attention from the slide. As the duration of attention could affect the cognitive load on working memory [19], with the tutee agent took some attention, it could be a positive way in lowering cognitive load of participants. Regarding the design of dialogue between tutor and the tutee agent, our participants gave their opinion. User6 commented on how the tutee agent can provides “break” from teacher/tutor:

“Well there is several things, first one is that the animated tutee agent summarize the content by parts make it easier to follow. And then, it provides a dialog so it’s not the same speakers all the time, probably it’s like giving a little bit of a break from the teacher...”

This particular comment is related to how we placed the tutee agent’s utterances at certain time in the videos, specifically at the end of each slide explanation. Participant felt that the tutee agent can give them time to break from the tutor’s explanation. As the participants cognitive load getting higher from long constant watch of the learning material presentation, we suggest that the tutee agent here can also help the students in decreasing their cognitive load as in learning segmentation [20]. Most of the participants who preferred the dialogue videos felt that the tutee agent’s summarizations help them in understanding the tutor explanation. User7 expressed how the dialogue videos make the learning more enjoyable:

“The dialogue version is more interesting and make me more enjoy to watch the video also. The animated person assistant also helps us to highlight what is the important content that the lecturer want to teach...”

Similar with User7 expressed, User11 has similar opinion:

“The animated agent help to summarize what the lecturer says, so I can memorize better about the key point in the lecture, each of the subtopics, or sub-lecture (subtopic) of the whole lecture...”

Based on these subjective comments from our participants, the combination of the right timing for utterance and the scaffolding in term of summarization by the pedagogical agent can help student in learning while watching lecture video, particularly in giving back their focus to the learning materials and affording them in lowering cognitive load.

7 Conclusion

Previous studies have explored strategies to improve lecture video by adding reducing cognitive load of the students by means of learning segmentation and breaks, adding activities such as using in-video quizzes and making annotation, and by incorporating pedagogical agents such as virtual tutor and virtual tutee. Using our tool, a teacher or course designer can design an interaction between tutor and the tutee agent in the monologue video as if they are making a dialogue. In this study we proposed our novel method to increase student attention to the lecture video by adding a tutee agent inside lecture videos.

Our statistical test from learning gain reported no significant effect of the tutee agent presence in dialogue video. On the eye fixation data captured using eye-tracker we concluded that the tutee agent existence in the dialogue video has some effects to the attention of the student. Observation to the data spread visualization on the average eye fixation frequency revealed that participants have more frequent eye fixations and shorter duration to slide area when watching dialogue videos. These results suggest that the tutee agent has some effect in increasing participants attention to the learning material while at the same time make their eye fixation duration shorter to it. When we asked participants on their opinion to the dialogue video, they confirmed that the tutee agent had effect on their attention to the video and learning material itself. This study still has limitation as it need more data from other sources to reveal more on the effect of dialogue video with the tutee agent by triangulation, particularly to the learning gain which will be also a future direction for our next investigation with larger participants. Another limitation that we note based on our participants' comment is how the tutee agent's summarization affect them. Whether the presence of the tutee agent that utter the summary, or the summarization itself has more effect to the participants need to be investigated more. We believe that by utilizing our prototype tool, further study about an effective way to design the tutee agent interaction in a monologue video, and how to design a lecture video which can be delivered either way in monologue or dialogue style are interesting.

References

1. Zhang, D., Zhou, L., Briggs, R.O., Nunamaker, J.F.: Instructional video in e-learning: assessing the impact of interactive video on learning effectiveness. *Inf. Manag.* **43**, 15–27 (2006). <https://doi.org/10.1016/j.im.2005.01.004>
2. Brame, C.J.: Effective educational videos: principles and guidelines for maximizing student learning from video content. *CBE Life Sci. Educ.* (2016). <https://doi.org/10.1187/cbe.16-03-0125>

3. Jing, H.G., Szpunar, K.K., Schacter, D.L.: Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *J. Exp. Psychol. Appl.* (2016). <https://doi.org/10.1037/xap0000087>
4. Mayer, R. (ed.): *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, Cambridge (2014)
5. Ilioudi, C., Giannakos, M.N., Chorianopoulos, K.: Investigating differences among the commonly used video lecture styles. In: *CEUR Workshop Proceedings*, vol. 983, pp. 21–26 (2013)
6. Chi, M.T.H., Kang, S., Yaghmourian, D.L.: Why students learn more from dialogue-than monologue-videos: analyses of peer interactions. *J. Learn. Sci.* **26**, 10–50 (2017). <https://doi.org/10.1080/10508406.2016.1204546>
7. Cohen, P.A., Kulik, J.A., Kulik, C.C., Cohen, P.A.: Educational outcomes of tutoring: a meta-analysis of findings. *Am. Educ. Res. Assoc.* **19**, 237–248 (1982)
8. Clark, R.E., Choi, S.: Five design principles for experiments on the effects of animated pedagogical agents. *J. Educ. Comput. Res.* **32**, 209–225 (2005). <https://doi.org/10.2190/7LRM-3BR2-44GW-9QQY>
9. Hayashi, Y.: Influence of social communication skills on collaborative learning with a pedagogical agent: Investigation based on the autism-spectrum quotient. In: *HAI 2015 - Proceedings of the 3rd International Conference on Human-Agent Interaction*, pp. 135–138 (2015). <https://doi.org/10.1145/2814940.2814946>
10. Craig, S.D., Gholson, B., Ventura, M., Graesser, A.C.: Overhearing dialogues and monologues in virtual tutoring sessions: effects on questioning and vicarious learning. *Int. J. Artif. Intell. Educ.* **11**, 242–253 (2000)
11. Park, S.W., Kim, C.: The effects of a virtual tutee system on academic reading engagement in a college classroom. *Educ. Tech. Res. Dev.* **64**(2), 195–218 (2015). <https://doi.org/10.1007/s11423-015-9416-3>
12. Shaw, E., Johnson, W.L., Ganeshan, R.: Pedagogical agents on the Web. In: *Proceedings of the International Conference on Autonomous Agents*, pp. 283–290 (1999). <https://doi.org/10.1145/301136.301210>
13. Kassner, M., Patera, W., Bulling, A.: Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction (2014). <https://doi.org/10.1145/2638728.2641695>
14. Liu, H., Heynderickx, I.: Visual attention in objective image quality assessment: based on eye-tracking data. *IEEE Trans. Circuits Syst. Video Technol.* **21**, 971–982 (2011). <https://doi.org/10.1109/TCSVT.2011.2133770>
15. Blokzijl, W., Andeweg, B.: The effects of text slide format and presentational quality on learning in college lectures. In: *IPCC 2005. Proceedings. International Professional Communication Conference, 2005, Limerick, Ireland*, pp. 288–299. IEEE (2005)
16. Driscoll, D.M., Craig, S.D., Gholson, B., Ventura, M., Hu, X., Graesser, A.C.: Vicarious learning: effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *J. Educ. Comput. Res.* **29**, 431–450 (2005). <https://doi.org/10.2190/q8cm-fh7l-6hjudt9w>
17. ELAN (Version 5.9). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive (2020)
18. Mayer, R.E., Dow, G.T., Mayer, S.: Multimedia learning in an interactive self-explaining environment: what works in the design of agent-based microworlds? *J. Educ. Psychol.* **95**, 806–812 (2003)
19. Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., Camos, V.: Time and cognitive load in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **33**, 570–585 (2007). <https://doi.org/10.1037/0278-7393.33.3.570>
20. Biard, N., Cojean, S., Jamet, E.: Effects of segmentation and pacing on procedural learning by video. *Comput. Hum. Behav.* (2018). <https://doi.org/10.1016/j.chb.2017.12.002>