LEARNING BY WATCHING TUTORIAL VIDEOS: IMPACTS OF DIALOGUE AND COLLABORATION

by © Xingbang Chen (Thesis) submitted

to the School of Graduate Studies in partial fulfillment of the

requirements for the degree of

Master of Education, Faculty of Education

Memorial University of Newfoundland

August 2022

St. John's Newfoundland and Labrador

Abstract

We are in a digital age when thousands of tutorial videos are available online. These videos are popular learning sources. However, we still know little about what makes a video powerful, and how to learn better by watching videos. A few studies show that observers learn better when collaboratively watching dialogue videos (recordings of conversations between a tutor and a tutee). However, the superiority of dialogue videos over monologue videos (recordings of a tutor's monologues) was not consistently shown in the literature. In addition, the advantage of collaboration over solo appeared only when observers watched dialogue videos. Those studies were limited to college student populations in English-speaking countries. The current study used a two-way between-subject design with pre- to posttest measures. It compared the learning outcomes of 60 students who watched tutorial dialogue and monologue videos alone and in pairs at a junior high school in China. Findings show that observers watching the dialogue video in pairs presented the most considerable learning effect and motivation. The results revealed a strong interaction effect between the dialogue video and collaboration. That is, only collaborative observers learned more from the dialogue video than from the monologue, and collaborative observers outperformed solo observers exclusively in dialogue observation.

Keywords: collaboration, learning from observing, tutorial dialogue, tutorial videos, vicarious learning

General Summary

The study compared the learning outcomes of 60 junior high Chinese students who learned a psychology topic in four conditions: watching a dialogue video (recording of conversations between a tutor and tutee) in pair, a dialogue video alone, a monologue video (recording of a tutor's monologues) in pair, a monologue video alone. The results showed that students learned more from the dialogue video than the monologue only in collaborative conditions. In addition, collaborative observers' learning gains exceeded solo observers' exclusively in dialogue conditions. The students who watched the dialogue video in pairs had the highest learning gains and motivation. There is a need for future research to explain the effectiveness of watching tutorial dialogues in pairs.

Acknowledgement

I want to express my appreciation to Dr. Tim Seifert, my supervisor, who gave me confidence and guided me through the research journey.

Also, this report cannot be here without the support of my family and friends. Thank you! Special thanks to my friend, Shalindra Senalankadhikara, who encouraged me along the way.

In addition, I appreciate all the help from Faculty of Education, School of Graduate Studies, Internationalization Office, and other departments of Memorial University during the pandemic.

Thanks to all the professors and instructors who have taught me, especially Dr. Cecile Badenhorst, who inspired me with freewriting techniques, making my thesis writing much easier.

Last but not least, thanks to everyone, especially my research assistant, who made this experiment happen.

Correspondence concerning this report should be addressed to Xingbang Chen, Faculty of Education, Memorial University of Newfoundland, xingbangc@mun.ca.

Contents

List of Tables

List of Figures

List of Appendices

Learning by Watching Tutorial Videos: Impacts of Dialogue and Collaboration

A tutor taught a student and recorded the tutoring session. When the recording was shown to other students, they had comparable learning gains to the tutee in the session. It may be surprising because we tend to think that without the tutors' direct feedback, observers' learning would not be as good as the tutee's. However, this happened in two experiments (Chi et al., 2008; Muldner et al., 2014), which provided evidence that learning by watching a tutoring recording (a tutorial dialogue video) can be as effective as being tutored.

The following chapter reviewed the literature on the effectiveness of observing a dialogue video and its alternatives. Effectiveness was represented by learning gains or outcomes in this report, which refers to knowledge or skills that learners obtained, such as their abilities to recall concepts or solve problems in a tutorial video. Then the present study comparing four observational learning approaches among a junior high Chinese population was presented together with results and discussions. This research added empirical evidence to the field of tutorial dialogues observation and gave both educators and learners more confidence in employing tutorial videos to save resources and promote learning.

Literature Review

This section first analyzed the pros and cons of learning by watching tutorial videos and being tutored. Then it reviewed researchers' attempts to improve the effectiveness of tutorial videos by creating dialogues in videos and allowing observers to collaborate while watching videos.

Tutorial Videos and Tutoring

With the help of the Internet and other innovations, tutorial videos have become one of the most promising teaching instruments (Andrist et al., 2014; Muller et al., 2007). Countless tutorial videos cover almost every aspect of human knowledge. Among these videos, one of the most common formats is monologue tutorials, in which only one person presents content. On popular online learning websites such as Udemy and Coursera, you probably will not find any style other than one instructor's monologues. In this report, tutorial videos refer to the videos that are created to help observers learn. One prominent advantage of video pedagogy is its economy. Once a tutorial video is created, it can be used unlimited times by innumerable learners.

Learning by watching videos is one kind of observational learning (both watching and observing include receiving audio and visual information in this report). According to Bandura's (1986) social learning theory, new behaviours can be acquired by observing and imitating others. In the 1960s, Bandura did a series of experiments that revealed the mechanisms of children's observational learning. In those experiments, children who observed adult models treating Bobo dolls aggressively in the same room (Bandura et al., 1961) or through videos (Bandura et al., 1963) produced more aggressive behaviours afterwards when they were presented with Bobo dolls, compared with children who observed nonaggressive behaviours and children in the control group who observed nothing. The experiment showed that children could learn adults' physical and verbal behaviours by simply observing them without being physically involved. Watching videos can be an effective way of learning.

Despite its pervasiveness, watching tutorial videos have a limitation—lack of interaction. When learners watch videos, they cannot talk to instructors in the videos and get feedback as they do in classrooms. The lack of interaction often impedes learning (Chi et al., 2001; Chi et al.,

2008; Schober & Clark, 1989). Tutoring, which allows interaction, is a highly effective pedagogy (Bloom, 1984; Chi et al., 2008; Graesser & McNamara, 2010; Muldner et al., 2014). During a tutoring session, tutors can receive feedback from tutees and thus know their cognitive levels, with which they can reach a "common linguistic and conceptual ground" (Geertshuis et al., 2021) with learners. On this "common ground", tutors can articulate better explanations that can be easily understood and thus foster learning. Another prominent feature of tutoring is frequent interaction between tutors and tutees. Research showed that interaction could improve learning effectiveness (Chi et al., 2001; Chi et al., 2008; Schober & Clark, 1989). Learning effects, gains, and outcomes in this report refer to cognitive competence and achievements (rather than emotions or other aspects), measured by various tests on learning subjects in different experiments.

However, human tutoring is not scalable (Chi et al., 2008; Geertshuis et al., 2021). That is, only a few learners can participate in a tutoring session and benefit from it. It is not cost-effective to assign a tutor to every individual learner. Learners usually need to pay a considerable amount of money for a tutoring session. Not every learner can afford it. Therefore, it creates an inequity among learners when only a few receive the benefits of the effective pedagogy, tutoring. Two factors—effectiveness and scalability—should be considered in deciding whether to adopt a pedagogical approach (Geertshuis et al., 2021).

Watching tutorial videos is scalable but not interactive, while being tutored is interactive but not scalable. Is there a method that has both advantages? Vygotsky's (1978) social constructivism theory points in a direction. Vygotsky identified two stages of learners' development: actual development and potential development. At the actual development stage, learners can understand materials and solve problems independently. By contrast, they need

guidance of a teacher or collaboration with a peer to understand materials and solve problems at the potential development stage. Vygotsky coined a term, "zone of proximal development", for the potential development stage, which is the moment when learning happens. If an observer is allowed to interact with a peer other than a tutor while watching monologue videos (a tutor's monologues), will they learn better? One study investigated the effectiveness of watching monologue videos in pairs (collaborating on a task while watching videos). In Craig et al.'s (2009) study, 67 college students watched monologue videos about rotational kinematics. It was found that learning outcomes from watching the monologue videos in pairs and alone were not significantly different, albeit the pair observers were engaged in discussion and collaboration to solve problems. More studies are needed to determine the impact of interaction in monologue video observation.

It should be noted that, among fifteen studies on tutorial videos that I reviewed, six of them employed a virtual tutor—a computer-controlled animated agent in an intelligent tutoring system—in their dialogue videos (Craig et al., 2000; Craig et al., 2004; Craig et al., 2006; Craig et al., 2009; Driscoll et al., 2003; Gholson et al., 2009). Little evidence proved that virtual tutoring produced comparable learning effects as human tutoring (Craig et al., 2006). However, their findings are similar to the research involving human tutors, as presented in the following sections.

Tutorial Dialogues

Most tutorial videos are monologues. If videos are recorded as dialogues, that is, filmed tutoring sessions, will observers learn better from watching dialogue videos? *Dialogues* mentioned in this report refer to tutorial dialogue videos, and *monologues* refer to tutorial monologue videos.

Dialogues vs. Tutoring

Craig et al. (2004) tackled this question in two experiments, where 120 and 110 university students studied computer literacy. Both experiments showed that students' learning effect size in virtual tutoring sessions was larger than in dialogue-observing groups. Here students interacted with a computer tutor by typing responses in a dialogue box in virtual tutoring sessions. In dialogue-observing groups, students watched videos of those virtual tutoring sessions. Chi et al. (2008) also found that watching dialogue videos alone was not as effective as being tutored. In their study, 70 undergraduates studied quantitative kinematics by interacting with a human tutor, watching a videotape of a human tutoring session either collaboratively or alone, collaborating with a peer, and studying alone. In an earlier study (Schober &. Clark, 1989), observers only received audio information without visual input, referred to as *overhearing*. Twenty students participated in tutoring sessions, and 40 students overheard those sessions. The 20 students in tutoring sessions worked in pairs, in which one tutored the other how to arrange tangram figures. The study showed that students who were tutored completed tangram tasks more accurately than overhearers. In sum, these studies indicated that observing tutorial dialogues was not as effective as being tutored.

Dialogues vs. Monologues

Though not comparable to being tutored, watching dialogues yielded larger learning outcomes than watching monologues in some studies (Driscoll et al., 2003; Fox Tree, 1999; Gholson et al., 2009; Muller et al., 2007). Two contributory features of dialogues were identified. However, the findings in favour of dialogues were inconsistent.

Superior Dialogues. Some researchers found that observers learned more from watching dialogues than monologues. In Craig et al.'s (2000) study, 48 university students studied

computer literacy by watching tutorial videos, in which both the tutor and the tutee were virtual agents from an intelligent tutoring system. In monologue videos, the virtual tutor presented all the information in a monologue style. The identical information was delivered through conversational exchanges between a virtual tutor and a virtual tutee in dialogue videos. Those who watched the dialogues asked the experimenter more questions in a post-observation task than those who watched the monologues, which was a sign of deeper reasoning that promoted learning (Craig et al., 2000). No learning outcomes were measured in this study. Gholson et al. (2009) replicated the study among 8, 9, 10, 11 graders, who studied computer literacy and Newtonian physics in the experiments. They compared pre-to-posttest learning gains among dialogue, monologue, and virtual tutoring conditions. It was found that the dialogue group had more gains than the monologue, and the gains in the tutoring condition were not significantly different from the monologue condition.

Two earlier studies compared the learning outcomes of students who overheard (audio only) dialogues and monologues. Fox Tree (1999) chose the same topic as the one in Schober and Clark's (1989) study: arranging tangram figures. One hundred and sixty-seven university students participated in the study, and those who overheard dialogues performed better in the task than those overhearing monologues. In Experiment 1 of Driscoll et al.'s (2003) study, 48 university students studied computer literacy. Learners who overheard dialogues wrote more content than those who overheard monologues. The dialogues were the conversations between a virtual tutor and a virtual tutee.

Contributory Components in Dialogues. What contributory components rendered dialogue observation superior to monologue in the studies above? Some researchers (Craig et al., 2006; Driscoll et al., 2003; Gholson et al., 2009) speculated that it was due to the deep-level

reasoning questions (hereafter *deep questions*) embedded in dialogues, which were absent in monologues. Deep questions ask about causes, processes, or consequences that scaffold learning and guide cognition. Examples are "Why…?", "How…?", and "What happens when…?" (Gholson et al., 2009). With this hypothesis, they included the same deep questions in both video formats. It turned out that the differences in learning outcomes were not significant between monologue and dialogue observation when identical deep questions were included in both. For example, in Craig et al.'s (2006) study, 140 undergraduates studied computer literacy. The pre-to-posttest learning gains were similar in the deep question dialogues and the deep question monologues conditions, higher than dialogues without deep questions. Gains in the virtual tutoring group were the same as the condition of dialogues without deep questions (deep questions dialogues = deep questions monologues > tutoring = dialogues). Similar findings were identified in Experiment 2 of Driscoll et al.'s (2003) study, in which 96 university students studied computer literacy. Learners who overheard dialogues with a virtual tutee's deep questions wrote more relevant content than those who overheard dialogues with the virtual tutee's shallow questions or comments or those overhearing monologues.

Misconceptions were another feature of dialogues, in which tutees presented wrong perceptions and experienced conceptual changes during tutoring sessions. In Muller et al.'s (2007) study, second-year physics class students studied quantum mechanics. Students who watched a dialogue video including a tutee's misconceptions had greater learning gains than those who watched monologues without misconceptions. In another study, Muller et al. (2008) introduced identical misconceptions from dialogues into monologues and found learning gains were equivalent, higher than monologues without misconceptions. Based on these studies, Lee (2019) balanced both misconceptions and deep questions in dialogue and monologue videos

about python programming, and again, the learning outcomes of 77 undergraduates in both conditions were not significantly different. However, in a post hoc analysis of Muldner et al.'s (2014) study, Chi et al. (2017) did not find strong correlations between observers' learning gains and deep questions or misconceptions. Nevertheless, with the evidence from experiments manipulating deep questions and misconceptions, it is hard to deny the impacts of these two variables on observers' learning.

Null Effects for Dialogues. Not all studies displayed superior dialogues. Null effects were found in two studies. In Cox et al.'s (1999) study, 54 undergraduates studied sentence parsing by watching dialogue and monologue videos. Pre-to-posttest learning gains were equivalent in the two conditions.

Most studies were conducted in lab settings. Cooper et al. (2018) moved the research into classrooms, where 280 university students took a physiology course. A Likert-scale survey question showed that nearly 60% of students preferred instructor-only (monologue) videos. In contrast with lab findings, students with a low grade point average (GPA) performed better in physiology quizzes after watching monologue videos compared with dialogues. For students with a median or higher GPA, dialogue and monologue videos did not significantly affect students' performance in the quizzes. However, as Cooper et al. (2018) noted, there were more uncontrollable factors in classrooms than in labs, such as other knowledge sources and distractions during video watching. Therefore, they did not reject the benefits of dialogue videos and suggested future research on this topic.

As mentioned, when two beneficial components—deep questions or misconceptions—from dialogues were copied into monologues, learning outcomes were equivalent in both formats. However, as Lee and Muldner (2020) believed, it is premature to conclude that delivery

format does not matter based on a handful of studies. There might be other merits of dialogue videos to be discovered, such as shared terminology between tutors and learners, learning skills presented by tutees in the videos and modelled by observers, improved self-efficacy of observers, motivations of observers from watching rewards and punishments in the videos, and close identities between observers and tutees (Geertshuis et al., 2021). All these factors may lead to different findings. Therefore, individual observation of dialogue and monologue videos requires further investigation, especially in less-explored younger populations.

Collaborative Observation

As mentioned in the section *Dialogues vs. Tutoring*, watching dialogue videos is not as effective as being tutored. Two exceptions were found in Craig et al.'s (2006) and Gholson et al.'s (2009) studies, where learners watching dialogue videos with deep questions had greater learning gains than tutees interacting with a virtual tutor. Is there any way to enhance the effectiveness of dialogue observation? In Craig et al.'s (2009) study, collaboration had a null effect on monologue observation. By contrast, a number of studies suggested collaboration fostered dialogue observation.

Collaboration Promoted Dialogue Observation

Three studies indicated that watching dialogues in pairs was more effective than watching dialogues alone and could be as effective as human tutoring. Chi et al. (2008) speculated that the inconsistent findings of the dialogue-monologue comparison resulted from uncontrolled active levels of observers. To trigger observers' active states, Chi et al. (2008) allowed some learners to collaborate while watching dialogue videos. Seventy undergraduates and one experienced tutor participated in the study, and the target domain was quantitative kinematics. Learners who watched dialogues in pairs had comparable learning gains to those in tutoring, greater than other

conditions (watching dialogues alone, collaborating, and studying alone). A similar finding was revealed in Muldner et al.'s (2014) study, where college students watching dialogue videos in pairs had equivalent learning gains as being tutored by human teachers. Craig et al. (2004) found that watching dialogues in pairs produced a larger learning effect than solo dialogue observation, but less than virtual tutoring. Craig et al. (2004) speculated that it was due to few conversation exchanges between dialogue observers. Otherwise, they would learn as much as those who received virtual tutoring. The conversations in Craig et al.'s (2004) experiment were less frequent than those in Chi et al.'s (2008) study.

Dialogue vs. Monologue Under Collaboration

Two studies revealed that watching dialogues in pairs also yielded a larger learning effect than watching monologues in pairs. In Craig et al.'s (2009) study, 67 college students studied rotational kinematics. Students who watched dialogue videos in pairs had greater learning gains than those who watched monologues in pairs. The same finding was in Muldner et al.'s (2014) study, Experiment 1, where 50 university students studied diffusion (particle movement). The collaborative dialogue group had higher learning gains than the collaborative monologue group. It seems collaborative dialogue observation is more effective than collaborative monologue observation.

However, not all studies followed this pattern. In Experiment 2 of Muldner et al.'s (2014) study, 40 junior high students studied diffusion. Students watching dialogue videos in pairs had similar learning gains to those watching monologues in pairs, both less than the gains in the human tutoring condition. Therefore, further research on collaborative observation is still needed, especially among young populations.

Based on these findings, Geertshuis et al. (2021) commented, "…it does not explain why individuals and collaborating pairs appear to learn more from watching dialogues than from watching monologues. Nor does it explain the nature of the impact of collaboration on learning." To explain this, first, we had to inspect the video designs in the two studies (Craig et al., 2009; Muldner et al., 2014). Deep questions and misconceptions were not controlled between the dialogues and the monologues. Therefore, superior dialogues might result from these beneficial features (deep questions and misconceptions) in the dialogue videos rather than observers' collaboration. Similarly, Lee and Muldner (2020) noted that "Since these features improve learning, it is an open question as to whether a dialogue format would improve observer learning from constructive activities while watching the video, over a monologue format, once these the content of the videos is equalized". Therefore, to answer Geertshuis et al.'s (2021) question and determine whether video formats matter in collaborative settings, future research is needed to compare collaborative dialogue and monologue observation. Meanwhile, factors such as deep questions and misconceptions have to be equalized across the two formats.

Research Questions

In sum, the domain of most studies was natural sciences. All experiments reviewed in this report except two (Gholson et al., 2009; Muldner et al., 2014) were conducted among university populations. In addition, all studies reviewed were set in English-speaking countries. Therefore, I conducted my research in a junior high Chinese population. Though it was found that solo dialogue observation was superior to solo monologue observation in four studies (Driscoll et al., 2003; Fox Tree, 1999; Gholson et al., 2009; Muller et al., 2007), this advantage disappeared when deep questions or misconceptions were controlled across the two video formats (Craig et al., 2006; Driscoll et al., 2003; Lee, 2019; Muller et al., 2008). Without controlling the deep

questions and misconceptions, two studies showed college students watching dialogues in pairs learned better than those watching monologues in pairs (Craig et al., 2009; Muldner et al., 2014), but Experiment 2 in Muldner et al.' (2014) report revealed a null effect among junior high school students. Because when deep questions and misconceptions were controlled, the experiments of solo dialogue-monologue comparison were scarce, and there was no study (to my knowledge) comparing pair dialogue-monologue conditions, the first research question is to be answered:

How do tutorial video formats impact observers' learning outcomes in solo and pair conditions when deep questions and misconceptions are controlled across a dialogue and a monologue?

When watching dialogues, college students who collaborated performed better than those who worked solo (Craig et al., 2004; Chi et al., 2008). In contrast, another college population who watched a monologue in pairs had similar learning gains to those who watched the monologue alone in Craig et al.'s (2009) study. Due to the few studies and divergent effects of collaboration, the second research question is to be answered:

How does collaboration influence observers' learning outcomes from watching tutorial dialogues and monologues?

Methodology

Participants

The participants (N = 64; 32 females, gender information provided by participants) were from a junior high school in a rural region of a central province in China. As reported by the school, the participants had lower academic achievements than their counterparts from urban regions in that city. With permission from the school, a research assistant visited two grade-eight classes and invited students to participate in this study after obtaining consent from the students and their parents. All participants were Chinese and spoke Mandarin. They were all grade-eight students, with ages from 13 to 15 (M = 13.7) years old. The participants' academic performance in a recent mid-term exam was obtained, including the scores of six subjects (Chinese, English, Ethics and Law, History, Math, and Physics), which was used to divide them into low and high achievers in later analysis.

Materials

Pre- and Posttest

The pre- and posttest consisted of the same fifteen multiple-choice questions with one or more correct choices. Each question was scored out of 10, and the total possible score was 150. I designed the test after filming the dialogue and monologue videos. The test was in Chinese but translated to English in this report. I created questions by selecting concepts that I deemed important from the videos. I invited several people to do pilot tests when they completed the tests before and after watching the videos. Previous experiments tested participants' ability to recall or apply the concepts from their tutorial videos. I did the same but with fewer concept application questions due to the nature of my content. Of the 15 questions, 14 required participants to recall the concepts directly stated in the videos, and one required them to apply a concept in their real lives (Question 12). For example, one of the questions was "Which statement is correct about the

hippocampus?" (see Appendix 1 for more questions). Participants completed the tests online, where the scores were calculated automatically, and the time spent on tests was recorded.

Muller et al. (2008) found that when misconceptions were introduced in both dialogue and monologue videos, there was no significant difference in their impacts on observers' learning. A subsequent experiment (Lee, 2019) drew a similar conclusion. To further investigate the effect of misconceptions in the two video formats, five misconceptions were included in the test, listed as follows:

1. *With frequent practice of the same movement, the hippocampus will regrow.*
2. *Short-term memory information is transferred from the cerebral cortex to the hippocampus and stored in the form of long-term memory.*
3. *Vitamin D can protect the biological enzymes in the hippocampus.*
4. *Vitamin D promotes the conversion of short-term memory into long-term memory.*
5. *Vitamin D will help the transmission of memory information.*

In the posttest, the first and second misconceptions above were presented as choices in Question 6, and the other three were the choices in Question 15 (see Appendix 1). Question 6 asked participants to choose the correct options. For each non-misconception option, they got 3 points. If participants chose Misconception 1 or 2, their choices were deemed to be mistakes, and they would not get any score on Question 6. Question 15 asked participants to choose the wrong options. Hence not selecting Misconception 3, 4, or 5 was considered a mistake. Participants got 3 points for each misconception they selected. However, they would lose all points on Question 15 if they picked the non-misconception option.

Questionnaire

A set of Likert-scale items was designed, similar to the ones in Cooper et al.'s (2018) study, to collect participants' opinions on the tutorial videos and the collaboration activity (see

Appendix 2). Every item included a statement, followed by 😁🙂😐☹️😣, representing five scales of agreement to the statement. 😁 stood for agreeing most, and 😣 for agreeing least. Participants also answered open-ended questions about reasons for their choices on Likert-scale items, and about their participation experience in general. Participants could not skip Likert-scale items, but the open-ended questions were optional. For example, Item 2 asked participants to rate the statement "I like the video" and write down the reasons for their choices. Participants responded to the questionnaire online.

Learning Task

To enable collaboration, I created a Learning Task that consisted of five questions (see Appendix 3). Participants needed to recall or apply the information from the videos. The tasks corresponded to the pre- and posttest. If participants completed the tasks correctly, they would be able to score the corresponding questions in the test. For example, the third question in Learning Task was, "Please list six activities you usually do. Which involves declarative memory, and which involves procedural memory? In what stages are these memories consolidated?" In the test, the twelfth question asked participants to choose activities that mainly depend on declarative memory. Moreover, in the thirteenth question, participants had to choose two correct options on memory types and their associated sleep stages. The tasks were printed on a piece of paper. The participants answered the questions in the Learning Task orally because it saved time, and I did not need to check their responses to the task.

Tutorial Videos

I recorded two tutorial videos, a dialogue and a monologue video, about memory and its relationship with diet and sleep. The topic was not part of the participants' school curriculum,

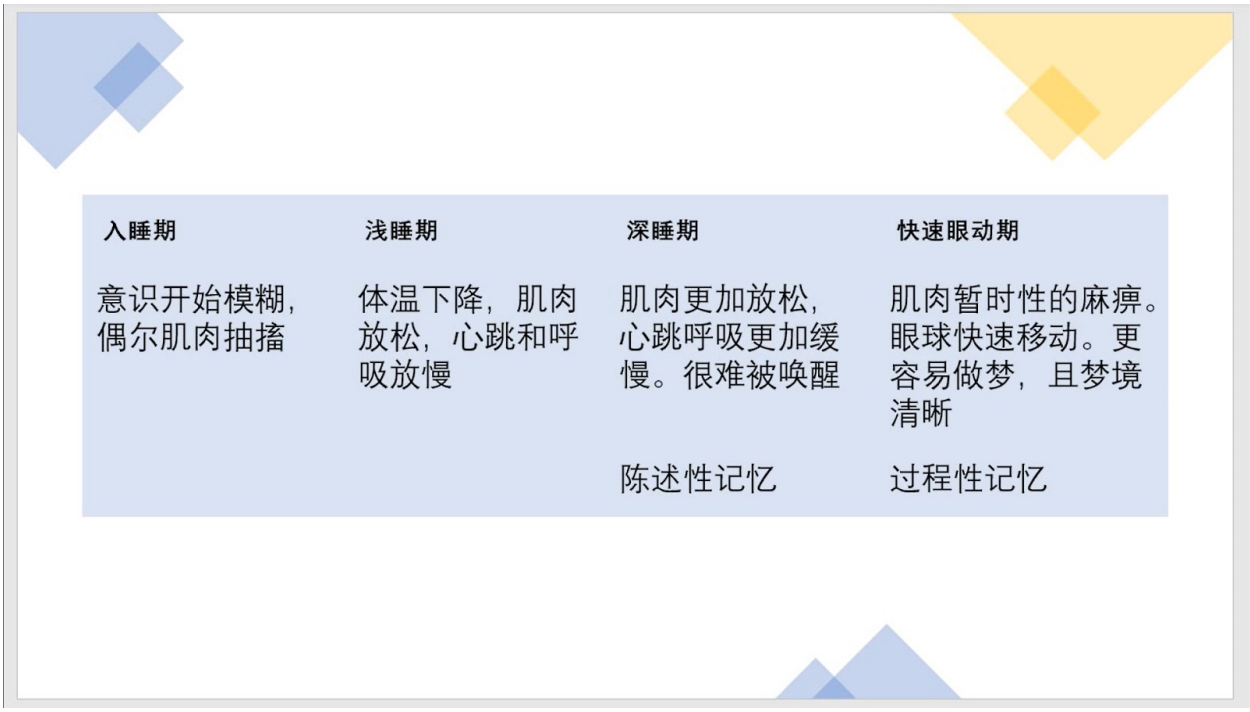
and the pretest scores also showed that participants had little prior knowledge of the topic (see Results).

I created one slideshow and presented it in both videos. The slides covered the following themes: the transition from short-term memory to long-term memory; the impact of Omega-3, Vitamin D, and alcohol on memory; the role of the hippocampus; declarative and procedural memory and their relationships with sleep stages.

The unscripted dialogue video is a recording of an online tutoring session with a 12-year-old grade-seven male student from another junior high school but the same town in China. During the session, I presented the slides to the tutee, and we discussed the topic. I narrated the topic, posed questions to the tutee, and corrected the tutee's misconceptions. I recorded the slides and our conversations with the software QQ Windows (Shenzhen Tencent Computer System Co., Ltd., 2022) on my Lenovo ThinkPad laptop. The length of the dialogue video is 25 minutes and 16 seconds.

I recorded the monologue video with the same laptop and software. I presented the identical content in the same order and details but in a monologue style (see Appendix 4 for dialogue and monologue excerpts). The slides were shown in the same order as in the dialogue video. Figure 1 is a screenshot from the monologue video, which also exists in the dialogue video. My voice and slides were recorded. The length of the monologue video is 18 minutes and 22 seconds.

I used a cursor to guide the tutee's attention on the slides, and cursor movement was visible in both slideshow videos. Both videos were original recordings without any editing. Neither the tutee nor I appeared in the video. Only our voice recordings were included.



Stage 1	**Stage 2**	**Deep Sleep**	**REM Sleep**
Brain activities start to slow down with periods of body twitches.	Body temperature drops and muscles relax. Heart rates and breathing slow down.	Muscles become more relaxed. Heart rates and breathing are slower. It is harder to wake up.	There is temporary muscle paralysis. Eyes move rapidly. Dreams are more common and vivid in this stage.
		Declarative Memory	Procedural Memory

Figure 1

A Screenshot from the Monologue Video (Top) and Translation of the Text (Bottom)

The previous studies indicated that deep questions (Craig et al., 2006; Driscoll et al., 2003; Gholson et al., 2009) and misconceptions (Muller et al., 2007; Muller et al., 2008) in videos could improve the observers' learning outcomes. The deep questions and misconceptions arose spontaneously during my interaction with the tutee. I controlled the two factors by first

recording the dialogue video, and then introducing the same deep questions and misconceptions into the monologue video, but only presented by me in the monologue. For example, one of the tutee's misconceptions is "Vitamin D can help short-term memory turn into long-term memory". I presented the misconception by saying, "Vitamin D cannot help short-term memory turn into long-term memory" in the monologue video. In the tutoring session, the tutee did not ask any question, and all the questions were posed by me. Therefore, I controlled deep questions by asking the same questions in both videos. In the dialogue video, the tutee responded to my questions. In the monologue video, I answered the questions myself.

Design and Procedures

The study used a two-way between-subject design with pre- to posttest measures. The participants were assigned to one of four groups, watching the dialogue video in pairs (Dialogue Pair), watching the dialogue video alone (Dialogue Solo), watching the monologue video in pairs (Monologue Pair), and watching the monologue video alone (Monologue Solo). I randomly assigned 32 female participants into the four groups: 10 in Dialogue Pair, 10 in Monologue Pair, 6 in Dialogue Solo, and 6 in Monologue Solo. I used the same method and randomly assigned 32 male participants. Hence there were 20 in Dialogue Pair, 20 in Monologue Pair, 12 in Dialogue Solo, and 12 in Monologue Solo. Then I randomly matched 10 same-gender pairs in Dialogue Pair and 10 same-gender pairs in Monologue Pair. The reason for the equal number of males and females in each group, and same-gender pairs is to avoid possible gender effects (Almasri et al., 2021; Bailey et al., 2020; Harskamp et al., 2008; Zhan et al., 2015).

The experiment took place at the junior high school where the participants came from. First, participants completed the online pretest within 10 minutes in a computer classroom. Then they watched the videos on desktop computers. In Dialogue Pair and Monologue Pair groups,

two participants sat in front of one monitor and were given the Learning Task on one piece of paper. In two Solo groups, each participant sat in front of one monitor and had the Learning Task sheet individually. All participants watched the videos simultaneously, controlled by the research assistant. Participants had no control of the videos because there were no headphones, and it would be distracting if they all played the videos at their own paces in the same room. After the videos, participants in Dialogue Pair and Monologue Pair discussed and completed Learning Task together. Participants in Solo groups completed the tasks alone. All participants did the Learning Task in 4 minutes. After that, each participant used one computer to take the posttest within 10 minutes. Participants in Pair groups took the test separately. Last, participants finished online questionnaires individually on the same computer where they took the posttest. The whole procedure took less than one hour.

Due to the seat limitation in the computer classroom and participants' available time in school, the experiment was conducted on two days. The participants in Monologue Pair and Monologue Solo joined in the experiment during a noon break. Two Dialogue groups followed the identical procedure at the same time the following day. To reduce the information that the Dialogue groups learned from the Monologue groups, I did not provide the participants with correct answers and final scores after they finished the pre- and posttest.

Results

The analysis was performed with the software JASP (JASP Team, 2021). Analysis of covariance (ANCOVA) was chosen for analyzing test scores because it could control the pretest

differences (Dimitrov & Rumrill, 2003), and it was often more powerful and appropriate than analysis of variance (ANOVA) on gain scores or repeated measures analysis of variance (RM ANOVA) in pre- to posttest designs (Dugard & Todman, 1995). In the current study, ANOVA on gain scores failed to control for the effect of pretest scores on posttest because gain scores were negatively correlated with pretest scores ($r = –.441$, $p < .001$). RM ANOVA was abandoned for another reason that the assumption of homogeneity of variance in pretest scores among four experimental groups was violated, $F(3, 56) = 5.27$, $p = .003$. To ensure that the assumptions of homogeneity of variances and normality were met, Levene's test and Q-Q plot of residues were checked in JASP for each ANCOVA. Alpha level was set at 0.05. A p-value less than 0.05 was considered a significant effect. When $0.05 < p < 0.1$, the finding was seen as marginally significant.

Learning Outcome

The participants used the same IP address to complete the tests in the computer classroom. However, the testing website restricted multiple requests from the same IP. As a result, the website asked random test takers to verify their submissions, which the participants were not instructed to do. The problem was solved the next day for the IP address being added to the safelist. Consequently, I did not receive complete test results from 4 participants (3 females) in the Monologue Pair group. They were missed completely at random. One female participant in the Monologue Solo group list missed the first-day experiment and joined Dialogue Solo the next day. Therefore, my analysis was based on the data from 60 participants: Dialogue Pair 20, Dialogue Solo 13, Monologue Pair 16, Monologue Solo 11.

Table 1

Means and Standard Deviations for Pre- and Posttest Scores, and Gains

Group		Pre			Post		Gain	
		N	M	SD	M	SD	M	SD
Dialogue	Pair	20	35.1	18.3	84.6	15.3	49.5	23.3
	Solo	13	33.0	21.8	54.4	26.0	21.4	21.3
Monologue	Pair	16	22.8	9.7	59.9	24.0	37.1	25.1
	Solo	11	33.8	16.9	61.8	14.9	28.0	23.2
Low achiever		30	28.3	17.5	60.7	22.5	32.4	24.8
High achiever		30	34.0	17.1	73.8	23.1	39.9	25.7
No-gain learner		7	42.6	24.2	37.1	26.3	–5.4	8.9
Plus-gain leaner		53	29.6	16.0	71.2	20.3	41.6	21.6

Pretest

A Kruskal-Wallis test was performed on the pretest scores because the data did not meet the assumption of homogeneity of variance among groups. The results showed that the pretest scores were not significantly different across the four groups, $H(3) = 4.391$, $p = .222$. As Table 1 presents, participants' average pretest scores ranged from 22.8 to 35.1 (The total possible score was 150). This suggested that participants had very limited knowledge about the learning content before watching the videos, and their pre-knowledge did not vary significantly across conditions.

Gender and Age

There were 29 females and 31 males in the analysis. In ANCOVA with gender as the independent variable, posttest scores as the dependent variable, and pretest scores as the covariant, there was no significant difference detected, $F(1, 57) = 2.11$, $p = .152$, $\eta_p^2 = .04$. There were 24 thirteen-year-old, 32 fourteen-year-old, and 4 fifteen-year-old participants. No

significant difference was detected in learning gains across age groups, according to ANCOVA with age as the independent variable, $F(2, 56) = 0.22$, $p = .807$, $\eta_p^2 = .01$.

Learning Effect Sizes

All four conditions had significant effect sizes on participants' pre- to posttest learning gains. Figure 2 shows the average scores changes with standard errors for the four groups. A paired sample t-test comparing pre- to posttest scores found the largest effect size in Dialogue Pair, $t(19) = 9.48$, $p < .001$, $d = 2.12$. The second largest effect size was in Monologue Pair $t(15) = 5.90$, $p < .001$, $d = 1.48$. It was followed by Monologue Solo $t(10) = 4.00$, $p = .003$, $d = 1.21$. Dialogue Solo had the least learning effects $t(12) = 3.62$, $p = .004$, $d = 1.00$.

Dialogue and Collaboration

ANCOVA was performed with posttest scores as the dependent variable and pretest scores as the covariate. There was no significant difference in participants' learning outcomes between Dialogue and Monologue groups, $F(1, 57) = 2.74$, $p = .104$, $\eta_p^2 = .05$, $d = 0.44$. On

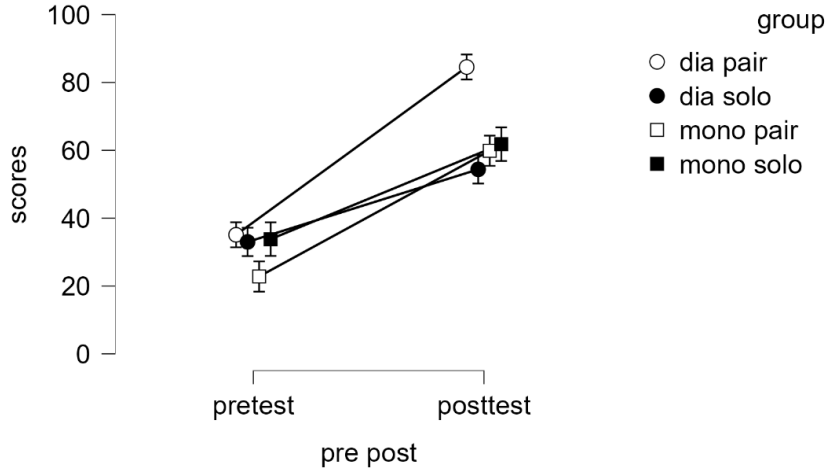


Figure 2

Mean Scores and Standard Errors from Pre- to Posttest for the Four Groups

Table 2

Planned Comparisons Among the Four Groups

Comparison	Mean Difference	SE	*df*	*t*	*p*	Cohen's *d*
dia pair vs. dia solo	29.6	7.2	55	4.13	< .001	1.47
dia pair vs. mono pair	21.2	7.0	55	3.04	0.004	1.08
dia solo vs. mono solo	–7.2	8.2	55	–0.88	0.385	–0.33
mono pair vs. mono solo	1.1	8.0	55	0.14	0.889	0.05

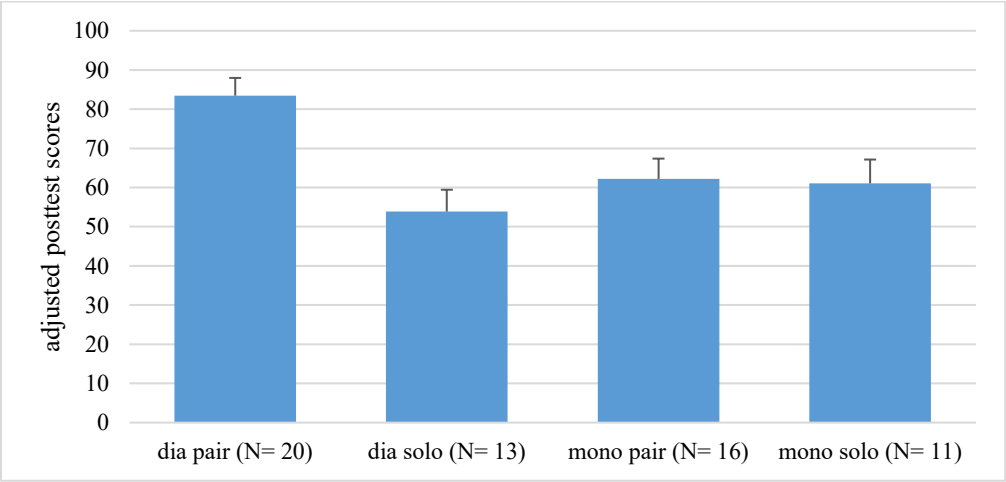


Figure 3

Adjusted Means and Standard Errors of Posttest Scores for Each Group

the other hand, participants performed significantly better in Pair groups than in Solo, $F(1, 57) = 9.32$, $p = .003$, $\eta_p^2 = .05$, $d = 0.77$. Among the four planned comparisons (see Table 2), only Dialogue Pair vs. Dialogue Solo ($p < .001$) and Dialogue Pair vs. Monologue Pair showed significant differences ($p = .004$). A Dunnett's post hoc test revealed that the learning gains of Dialogue Pair were also significantly larger than those of Monologue Solo ($t = 2.96$, $p = .013$),

proving that Dialogue Pair had the greatest learning outcomes among the four groups. Figure 3 illustrates the four groups' average posttest scores adjusted with pretest scores.

Low and High Achievers

The participants were ranked according to their total scores of the six subjects in a mid-term exam. The first half was labelled as high achievers and the other as low achievers, with 30 people for each half. The scores of two types of achievers can be found in Table 1. An independent t-test demonstrated that high achievers scored significantly higher than low achievers in the mid-term exam, $t(58) = 11.94$, $p < .001$, $d = 3.08$. The pretest scores of the two types were statistically equivalent, $t(58) = 1.27$, $p = .210$, $d = 0.33$. ANCOVA showed that higher achievers had a marginally greater learning gains than low achievers, $F(1, 57) = 3.79$, $p = .057$, $\eta_p^2 = .06$, $d = 0.50$. Low achievers who worked in pairs had a significant larger learning outcome than those completed the task solo, $F(1, 27) = 4.69$, $p = .039$, $\eta_p^2 = .15$, $d = 0.78$. The learning outcomes between low achievers watching the dialogue video and those watching the monologue were not significantly different, $F(1, 27) = 1.47$, $p = .237$, $\eta_p^2 = .05$, $d = 0.44$. Four planned comparisons (same as Table 2) revealed two significant contrasts: Dialogue Pair vs. Dialogue Solo ($d = 1.84$, $p < .001$) and Dialogue Pair vs. Monologue Pair ($d = 2.13$, $p = .008$).

High achievers in Pair groups had a marginally significantly larger learning outcome than those completed the task solo, $F(1, 27) = 3.11$, $p = .089$, $\eta_p^2 = .10$, $d = 0.68$. Same as low achievers, there were no significant difference in high achievers' learning outcomes between dialogue watching and monologue watching, $F(1, 27) = 0.96$, $p = .337$, $\eta_p^2 = .03$, $d = 0.38$. The four planned comparisons revealed the same two significant contrasts as in low achievers but with smaller effect sizes: Dialogue Pair vs. Dialogue Solo ($d = 1.53$, $p = .014$) and Dialogue Pair vs. Monologue Pair ($d = 0.96$, $p = .043$).

When split into low and high achievers, ANCOVA mirrored the pattern in the analysis of the whole sample. However, interventions had a larger effect on low achievers than high achievers, that is, low achievers were more sensitive to experimental treatments.

Chi et al. (2008) divided students into Good and Poor Observers based on a median split of their pretest scores after the students studied materials relevant to the pretest. In my study, grouping based on the pretest could not represent the participants' learning achievements because they did not study the content before the pretest, and their pretest scores were likely results of guessing.

No-gain Learners

Interestingly, more participants had zero or minus learning gains in Solo groups than in Pair groups. The number of zero or minus learning gains in each group was: Dialogue Pair 0, Dialogue Solo 3, Monologue Pair 1, Monologue Solo 3. The scores of no-gain learners can be found in Table 1. I speculated that zero or minus learning gains were due to a lack of effort when they took the tests, which could be indicated by the time they spent on tests. Some participants might rush through the tests because the scores would not affect their course grades.

My assumption was supported by ANOVA on time spent on tests. The time that participants spent on the pre- and posttest was recorded when they completed the tests online. I compared the 7 no-gains with the other participants. The result of their time spent on the pretest did not reveal any significant difference, $F(1, 58) < 0.01$, $p = .950$, $\eta_p^2 < .01$. However, for the posttest, the 7 participants spent marginally significantly less time on completing the posttest by

90.5 seconds on average, $d = –0.76$, $p = .063$. The average time for the other participants was 386.5 seconds.

The 7 participants' posttest scores might not reflect their learning outcomes because they probably did not spend enough time completing the posttest. Would an analysis without their data show a different result? When 7 participants with zero or minus learning gains were removed, ANCOVA with the pretest scores as covariant found a significant difference of conditions' effects on the posttest scores, $F(3, 48) = 4.80$, $p = .005$, $\eta_p^2 = .23$. The four planned comparisons as in Table 2 was run again and revealed only two significant comparisons: Dialogue Pair vs. Dialogue Solo ($d = 1.29$, $p = .002$) and Dialogue Pair vs. Monologue Pair ($d = 1.00$, $p = .007$). The finding was the same as that from the data containing the 7 participants.

Time and Test Scores

Does shorter time spent on tests predict lower scores? Correlation tests between time length and test scores revealed that pretest scores and time did not show strong relationships ($r = .28$, $p = .031$). However, the correlation between posttest scores and time was slightly moderate ($r = .55$, $p < .001$).

With the correlation between time and scores, I expected the analysis of time variances would show a similar pattern as the test scores. ANOVA on time spent on the pretest was performed, and Holm-Bonferroni sequential corrected post hoc testing revealed only one significant comparison: Dialogue Pair vs. Monologue Pair ($p = .027$, $d = 1.11$). A Kruskal-Wallis test was performed on time spent on the posttest because the data did not meet the assumptions of homogeneity of variance. Corrected by Holm-Bonferroni sequential method, a non-parametric post hoc test (Dunn) found two significant comparisons on time spent on posttest: Dialogue Pair

vs. Dialogue Solo ($p < .001$, $z = 4.69$); Dialogue Pair vs. Monologue Pair ($p = .008$, $z = 2.93$), which also appeared to be significant comparisons in the analysis on test scores. Figure 4 and Figure 5 illustrate the standard errors and means of time and scores respectively for each group.

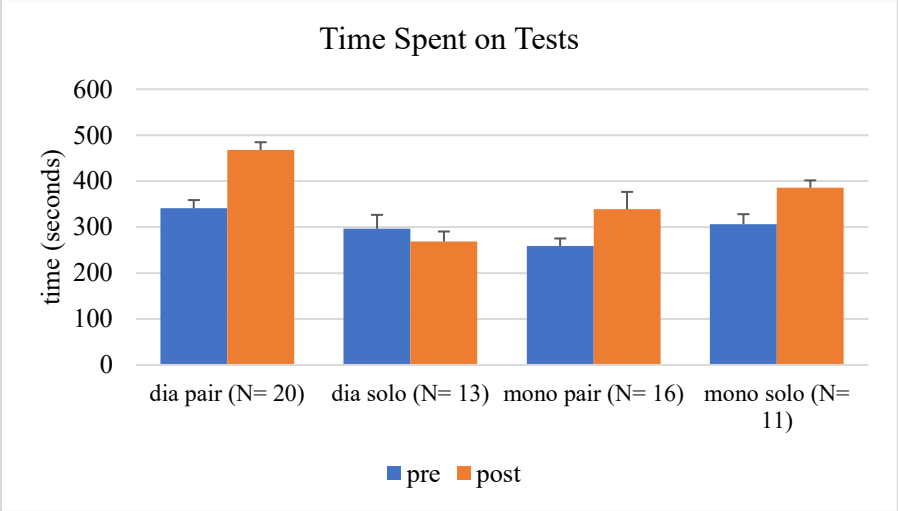


Figure 4

Standard Errors and Means of Time Spent on Tests for Each Group

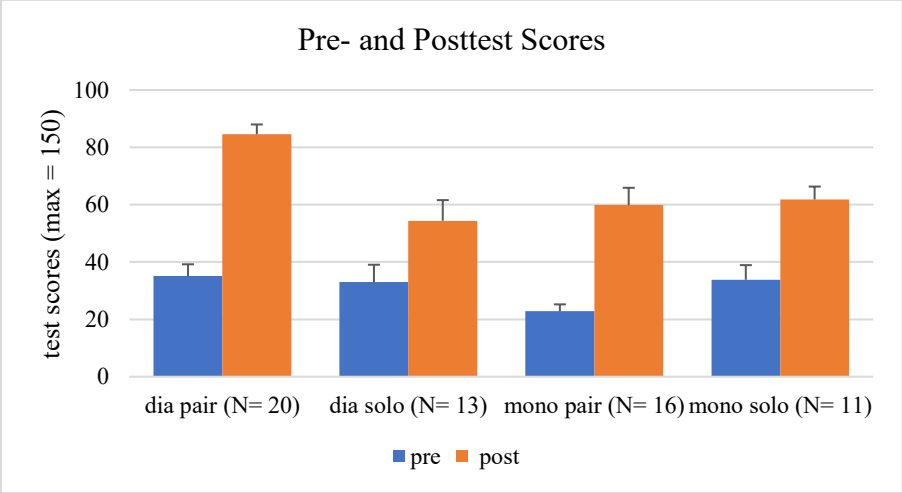


Figure 5

Standard Errors and Means of Test Scores for Each Group

However, ANOVA could not tell that the differences in time spent on the posttest were due to the four interventions because time spent on the pretest might influence time on the posttest. To control the pretest time's influence, I performed ANCOVA with time spent on the pretest as the covariant and found a strong effect of pretest time on posttest time, $F(1, 54) = 5.71$, $p = .020$, $\eta_p^2 = .10$. There was also a significant interaction between pretest time and observational conditions, $F(3, 51) = 4.48$, $p = .007$, $\eta_p^2 = .21$. Since the assumption of homogeneity of variances was violated, a Games-Howell post hoc test was performed, corrected with the Turkey method. The test revealed four significant comparisons: Dialogue Pair vs. Dialogue Solo ($t = 7.20$, $p < .001$), Dialogue Pair vs. Monologue Pair ($t = 3.25$, $p = .026$), Dialogue Pair vs. Monologue Solo ($t = 3. 52$, $p = 0.008$), Dialogue Solo vs. Monologue Solo ($t = –4.31$, $p = .002$). The result indicated that watching dialogues in pairs encouraged participants to spend more time on the posttest than other conditions. Time analysis (ANOVA) between low and high achievers did not show any significant differences on either pre- or posttest, suggesting that they spent a similar amount of time on two tests.

Time on the posttest could be another factor that influenced learning outcomes. ANCOVA with both pretest scores and time on the posttest as covariant was performed and revealed no significant difference among the four groups, $F(3, 54) = 1.93$, $p = .135$, $\eta_p^2 = .10$. Posttest time had a significant effect on posttest scores, $F(1, 54) = 7.19$, $p = .010$, $\eta_p^2 = .12$.

The Impact of Misconceptions

As discussed in the Methodology section, there were five misconceptions in the tests. I collapsed the two dialogue conditions and the two monologue groups into two categories, Dialogue and Monologue, and recorded the number of participants making mistakes on Misconception 1 in the posttest (see Table 3). Two of the four missing data in Monologue Pair

contained the two participants' responses to the misconceptions and were included in the following chi-square analysis. A chi-square test of independence was performed to examine the relationship between Dialogue/Monologue video and Right/Wrong choice on Misconception 1. The relationship between these variables was not significant, $\chi^2(1, N = 62) = 0.01, p = .923$.

To further investigate the impact of misconceptions on the two video formats, I controlled collaboration conditions and conducted two other chi-square tests in Pair and Solo conditions separately (see Table 3). The relationship between Dialogue Pair/Monologue Pair and Right/Wrong choice on Misconception 1 was not significant, $\chi^2(1, N = 38) = 2.46, p = .302$. Neither was the relationship between Dialogue Solo/Monologue Solo and Right/Wrong choice, $\chi^2(1, N = 24) = 1.70, p = .193$.

When I collapsed the four groups into two conditions, Pair and Solo, and performed a chi-square test of independence, there was no significant relationship between Pair/Solo observation and Right/Wrong choice on Misconception 1, $\chi^2(1, N = 62) = 0.26, p = .613$ (see Table 3). When I examined the relationships within Dialogue and Monologue separately (see Table 3), chi-square tests did not reveal any significant relationship either.

Twenty-four chi-square tests (see Appendix 5) were also performed on the other four misconceptions, and no significant relationship was found except Monologue Pair/Monologue Solo on Misconception 2, $\chi^2(1, N = 29) = 3.99, p = .046$ and Dialogue Pair/Dialogue Solo on Misconception 4, $\chi^2(1, N = 33) = 4.41, p = .036$. The results showed that misconceptions in the dialogue and the monologue conditions had similar influences on observers. So was the effect on Pair and Solo conditions.

Table 3

The Number of Participants Making Mistakes on Misconception 1 in the Posttest After Watching the Dialogue/Monologue Videos in Pairs/Solo

Mis 1	Right	Wrong	$\chi^2(1)$	*p*
Mono	24	5	0.01	0.923
Dia	27	6		
Mono Pair	14	4	1.06	0.302
Dia Pair	18	2		
Mono Solo	10	1	1.70	0.193
Dia Solo	9	4		
Pair	32	6	0.26	0.613
Solo	19	5		
Dia Pair	18	2	2.28	0.130
Dia Solo	9	4		
Mono Pair	14	4	0.83	0.360
Mono Solo	10	1		

Questionnaire Results

For unknown reasons, the responses from the three participants were not received (Dialogue Pair: 1, Monologue Pair: 1, Monologue Solo: 1). As a result, there were 61 responses collected.

Participants' responses to the questionnaire items with five choices were presented in Table 4. The item "I like the video" collected participants' attitudes towards the two videos. Due to the between-subject design, each participant shared their attitude to one of the videos. The result showed that participants rated more favourably on the dialogue video than the monologue (see Figure 6). All 7 no-gain learners chose 😄 for the videos they watched, except that one chose 😐 for the monologue video. Among participants who watched the dialogue video, 82.4% (including all no-gain learners who watched the dialogue video) selected 😄 under the statement

"the tutee in the video helped me learn". All participants in Pair groups (100%) responded to the statement, "I like watching the video and discussing the questions with my partner" with 😄. Even though most participants enjoyed the video learning experiences, when provided with the statement "I hope to interact with the teacher directly rather than watching a video", 82.0% (including all no-gain learners) chose 😄.

Table 4

Participants' Responses to the Questionnaire Items

Item \ Choice	😄	🙂	😐	😕	😖
2. I like this video. (Dialogue)	91.2	0.0	8.8	0.0	0.0
4. The student in the video was helpful to my watching and learning.	82.4	2.9	8.8	0.0	5.9
5. The student in the video interfered with my watching and learning.	20.6	5.9	29.4	2.9	41.2
6. I hope that there is no student' voice in the video.	23.5	2.9	29.4	2.9	41.2
8. I like this video. (Monologue)	66.7	11.1	18.5	0.0	3.7
10. The teacher in the video was helpful to my watching and learning.	98.4	1.6	0.0	0.0	0.0
11. I was very focused while watching the video.	90.2	3.3	6.6	0.0	0.0
13. Rather than watching a video, I hope to interact with the teacher directly.	82.0	1.6	8.2	0.0	8.2
16. I like to watch the video and discuss problems with my partner.	100.0	0.0	0.0	0.0	0.0

Note. All numbers are rounded percentages so the sums may not be 100%.

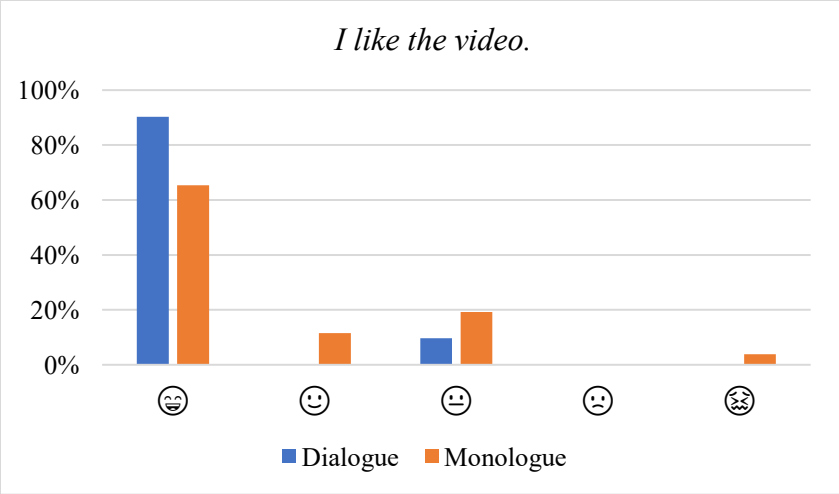


Figure 6

Participants' Responses to "I like the video."

Regarding the statement "The tutee in the video helped me learn", the questionnaire asked participants to explain their choices in text format. Regarding the responses from 31 participants, I created a code for each response and grouped the responses with the same code together. I compared the code and original responses to check the consistency. Then I created two categories, *advantages* and *disadvantages* of the dialogue video, and selected the codes into the two categories. Table 5 shows the result of this coding.

The questionnaire also asked participants why they liked watching the video and discussing the questions with partners. I used the same coding method and analyzed the responses from 38 participants. Only the advantages of collaboration were generated because all comments from participants were positive (see Table 6).

Table 5

Participant-perceived Advantages and Disadvantages of the Tutee's Presence in the Video

Advantages:

Code	Descriptions	Example quote
Helpful	The tutee in the video helped observers learn.	"His words helped me learn."
Interesting	The tutee's presence made the video interesting.	"The interaction between the tutor and the tutee made the class interesting."
Similar perspectives	The tutee and observers shared similar perspectives.	"Because I think the answers from the student in the video were similar to most of mine, and I could know my mistake through him."

Disadvantages:

Code	Descriptions	Example quote
Distraction	The tutee's voice distracted observers from watching and learning.	"The student' voice was abrupt, preventing me from viewing and remembering the content."
No influence	The tutee's presence had no impact on observers.	"His answer did not affect me."

Table 6

Participant-perceived Advantages of Collaboration

Code	Descriptions	Example quote
Interesting	Participants found collaboration and discussion interesting and engaging.	"I felt engaged in learning."
Better learning	Collaboration and discussion helped participants learn more and improve their retention.	"Because it helped me understand more."
Achievement	Participants obtained a sense of achievement after collaboration.	"Completing the tasks with my partner gave me a sense of achievement."
Emotion	Collaboration gave participants emotional comfort.	"I didn't feel lonely during collaboration."

Summary

Participants' responses to the questionnaire showed that more of them felt positive towards the dialogue video, and most participants believed that the tutee's voice in the video helped them learn. All participants rated their collaborative learning experience highly positive. However, participants' subjective feelings only partially reflected their learning outcomes in the pre- and posttest. The analysis suggested that the dialogue video helped participants learn better than monologue only in collaborative conditions, and collaboration fostered participants' learning more than solo exclusively in dialogue conditions. Dialogue Pair had the most considerable learning gains and spent the longest time on the posttest among the four groups.

Also, I controlled misconceptions across the two videos and included them as choices in the tests. The analysis of participants' responses in the posttest suggested that misconceptions had similar impacts on participants' learning regardless of video formats or collaborative modes.

Discussion

The present study found the superiority of the dialogue video over the monologue only in collaborative groups. Similarly, the advantage of collaboration over solo appeared exclusively in dialogue observation. These discoveries are consistent with most previous findings.

The Effects of Dialogue and Collaboration

First, the results on video formats replicate what was found in Craig et al.'s (2009) experiment and Muldner et al.'s (2014) Study 1. I controlled deep questions and misconceptions in both the dialogue and the monologue videos, while the other two did not. Nonetheless, I came to the same finding, learners who watched the dialogue video in pairs had higher learning gains than those who watched the monologue in pairs.

However, a short-term assessment in Craig et al.'s (2009) experiment yielded similar learning effects between Dialogue Pair and Monologue Pair. Craig et al. (2009) conjectured that students might rush through the short-term assessment due to their careless attitudes toward the short-term assessment and prioritization of the long-term measurement in that classroom-based experiment. Likewise, Study 2 in Muldner et al.'s (2014) report shows a null effect, where junior high students had similar learning gains between Dialogue Pair and Monologue Pair. Muldner et al. (2014) speculated that the benefits of collaboration might overshadow that of dialogues. This

is reasonable because compared with 4 minutes of collaboration in my experiment, students in Muldner et al.'s Study 2 worked in pairs for 36 minutes on average. My results are aligned with Muldner et al.'s belief that collaboratively watching dialogue videos can also benefit young populations.

Also, my results are consistent with the studies on deep questions and misconceptions (Craig et al., 2006; Lee, 2019; Muller et al., 2008). Video formats did not affect Solo observers' learning outcomes when video content, deep questions, and misconceptions were controlled. Additionally, the analysis of participants' responses to five misconceptions in the posttest illustrated that misconceptions in the dialogue and monologue videos had similar effects on observers' learning, either in pairs or solo.

The present results on collaboration mirrored the pattern in Chi et al.'s (2008) study: Learners who watched the dialogue video had larger learning outcomes when they worked in pairs compared with solo. Moreover, my study showed that collaboration did not have noticeable effects on Monologue groups, which was also found by Craig et al. (2009). In Craig et al.'s study, college students in Monologue Pair and Monologue Solo had similar learning outcomes. Craig et al. used long-term measures (26 days on average after interventions), but I tested participants' short-term retention right after the interventions.

In sum, the results show a strong interaction effect between video formats and collaboration: Effective learning depends on both the dialogue video and collaboration.

Explanations of the Interaction Effect

The present results were not expected because I assumed that collaboration would improve learning outcomes, and video formats would not make a difference. However, it was not

the case—only Dialogue Pair outperformed Dialogue Solo. Monologue Pair did not distinguish itself from Monologue Solo. Additionally, though Dialogue Solo had an equivalent impact as Monologue Solo, Dialogue Pair helped observers more than Monologue Pair. Why was this inconsistency? Previous studies provided explanations.

Chi et al. (2017) analyzed the data from Chi et al.'s (2008) and Muldner et al.'s (2014) studies and concluded that watching dialogue videos elicited more constructive and interactive behaviours from pair observers compared with watching monologues. Constructive behaviours are pair observers' substantive comments, which are statements relevant to the topic being taught. Interactive behaviours are pair observers' conversations on the same matter that involve at least one substantive comment. Similarly, Craig et al. (2009) found that compared with Monologue Pair, Dialogue Pair exhibited more constructive and interactive behaviours, such as actively engaged in discussing the topic, identifying knowledge discrepancies between pairs, explaining and solving the tasks during collaboration. Chi et al. (2017) also identified positive correlations between observers' constructive and interactive behaviours with their learning outcomes. With this explanation, Solo groups in my study had lower performance because they did not have a collaborative context to be constructive and interactive. Monologue Pair was not comparable to Dialogue Pair for the absence of the dialogue video's influence.

Why do dialogue videos make pair observers more constructive and interactive? Chi et al. (2017) provided two reasons. First, it was found that in Chi et al.'s (2008) and Muldner et al.'s (2014) studies, observers' constructive and interactive behaviours were positively correlated with those of videoed tutees, indicating that observers modelled tutees' behaviours. The modelling mechanism would be similar to how the children learned aggressive behaviours by watching videos in the Bobo doll experiment (Bandura et al., 1963). In my study, learners might also

model the tutor's behaviours because the tutee was not actively inquiring, but the tutor frequently generated questions to entice the tutee to think. Lee (2019) gave eye-tracking evidence that observers paid more attention to a tutor than a tutee in a dialogue video. Second, Chi et al. (2017) also discovered a positive correlation between conflict episodes in the dialogue videos and observers' constructive and interactive behaviours. Conflict episodes were interactions in the dialogue videos when tutees made a mistake, later corrected by tutors. Chi et al. (2017) postulated that tutees' mistakes motivated observers to avoid mistakes themselves and try harder, represented by more effortfully constructive and interactive behaviours during collaborative observation.

Participants' modelling behaviours and motivation can explain a finding in my study that low achievers were more sensitive to experimental treatments. That is, low achievers benefited more from collaborative dialogue observation, as revealed in the Results chapter. Based on the explanations above, maybe high achievers already possessed more collaborative skills or higher learning motivation compared with low achievers, and thus collaborative dialogue observation gave a smaller plus to high achieves' learning. However, the speculation that high achievers have higher learning motivations is not well supported because my results show that low and high achievers spent a similar amount of time on either pre- or posttest. Future research is needed to verify my findings on low and high achievers, which can help us understand more about the relationship between learners' differences and observational learning.

It should be noted that Chi et al. (2017) did not build any direct correlations between tutees' behaviours or conflict episodes and observers' learning gains. Researchers can manipulate these two factors in future experiments. For example, studies can compare observers' modelling behaviours and learning outcomes from the video with different amounts of tutee's

constructive and interactive behaviours, such as a dialogue video with more such behaviours, a dialogue with fewer these behaviours, and a monologue video. There may be other attributes of dialogue videos apart from tutees' behaviours and conflict episodes. Explorative research may discover potential advantages of dialogue videos.

Motivation

Motivation in this report refers to participants' willingness to engage in learning activities. Chi et al. (2017) regarded constructive and interactive behaviours during collaboration as indicators of learners' efforts. Differently, I used the time participants completed a test in my study as a sign of their engagement. The analysis of time spent by participants on the posttest mirrors that of learning outcomes: Dialogue Pair spent more time on the posttest and had higher scores than Dialogue Solo or Monologue Pair. The fact that Dialogue Pair spent the longest time in completing the posttest suggested that they were most motivated to try hard in the posttest. If the motivation is attributed to dialogue observation alone, Dialogue Solo should have also spent more time than monologue groups, which was not the case. Therefore, motivation to perform well in the posttest should arise from both the dialogue video and collaboration. Collaboration could trigger pair observers' motivation to compete because they might want to outperform their partners in the posttest.

However, it leaves a question: What is the relationship between motivation and learning outcomes? One possibility is that dialogue videos improve pair observers' motivation, and higher motivation leads to better learning outcomes. Or watching dialogue videos in pairs improves learning gains, which boosts motivation. Another alternative is that dialogue videos improve pair observers' motivation and learning outcomes, that is, motivation is a by-product. The current

study cannot answer this question. Further research is needed to explain the relationships among observational conditions, learning outcomes, and motivation.

So far, some evidence (Chi et al., 2008; Craig et al., 2009), including the current study, indicates that collaboration improves learning outcomes of dialogue observers but not monologue observers. This supports Kuhn's (2015) conclusion that cognitive collaboration will not always improve learning outcomes but depends on certain conditions. Observing dialogue videos seems to be one condition that evokes collaboration's benefits.

Advantages of Dialogues

The participants' responses in the questionnaire show an advantage of dialogue videos. When asked why they liked the tutee's presence in the video, participants noted that the tutee in the video helped them learn by making similar mistakes as they did, which contributed to their learning. This indicates that an advantage of a tutee in the tutorial video is offering a close perspective to observers'. Chi (2013) described this phenomenon with the term, *zone of proximal representational match*, borrowed from (Chi et al. 2017) Vygotsky's (1978) phrase, *zone of proximal development*. *Zone of proximal representational match* means that compared with a tutor and observers, there is less mismatch between a tutee in a dialogue video and observers who watch the video, if they are both novice learners. It allows the tutee and observers to share understandings and terminology of learning content, which contributes to observers' learning. Likewise, Geertshuis et al. (2021) reviewed the past works and summarized that observational learning outcomes would be better if a videoed tutee and observers had similar identities, abilities, resources, and preparation. This is why I chose the tutee of similar age and background as the participants in my study. However, the variance of videoed tutees in dialogue videos is a research topic to be investigated. How do tutees of different competence, ages, and backgrounds

in videos affect observational learning? Only one study was found relevant to the question. With a small sample size of 20, Chi et al. (2008) discovered that collaborative observers learned more from watching dialogues with tutees of higher competence.

Even though Solo participants who watched the dialogue and the monologue videos did not significantly differ in learning outcomes, we should not overlook the educational utility of dialogue videos. Without the dialogue video from the tutoring session, I would not obtain misconceptions from the tutee, which were appreciated by some observers, "The tutee's answers were similar to mine, so I could know my own mistakes by watching him." Also, the previous research proved the effect of misconceptions on fostering observers' learning. It is easier for instructors to collect genuine misconceptions from learners when producing tutorial dialogues than to conjecture artificial misconceptions themselves.

Differences in Collaboration and Videos

The collaboration style in this study is different from the previous ones (Chi et al., 2008; Craig et al., 2004; Muldner et al., 2014). In my study, participants did not manipulate the videos, but participants in the prior experiments controlled videos by stopping, rewinding, or fast-forwarding. Nevertheless, all studies show that participants who watched dialogues in pairs had superior learning outcomes than those who watched dialogues solo or monologues in pairs. This indicates that it is effective when collaboration happens either during or after dialogue observation.

Another variation is that my tutorial videos did not have any tutor's or tutee's appearances. That is, observers only heard conversations between the tutor and the tutee while watching the slideshow. Nevertheless, the absence of visual presence did not lead to different

conclusions. Lee (2019) compared tutorial monologues on python programming with and without a tutor's talking head and found no difference in college students' learning gains, cognitive loads, interests, and feelings of social presence from watching the two types of monologues. An opportunity for future investigation is to compare tutorial dialogues with and without a tutor's and a tutee's visual presence.

Limitations

Due to the nature of classroom research, my sample size is limited to 60 participants, which may lead to a question of its statistical power. However, the analysis with G*Power software (The G*Power Team, 2020) proved that this study is of high statistical power. An ANCOVA post hoc test with a 95% significance level, a sample size of 60, and an effect size of 1.28 generated a statistical power of 100%. The effect size of 1.28 is the average of the two comparisons in the Results chapter: Dialogue Pair vs. Dialogue Solo (d = 1.47) and Dialogue Pair vs. Monologue Pair (d = 1.08).

There is only one tutor and one tutee in my videos. Hence, experiments with other tutors or tutees may show different results if observers respond in various manners to different people in videos. Future research can recruit multiple tutors and tutees, as did in Mulder et al.'s (2014) study, to control specific tutors' or tutees' influence if resources are available. Also, with various tutors and tutees, there can be more than one dialogue and monologue video in an experiment, which will improve generalizability.

Some environmental factors (noises) might influence participants' learning because the experiment was conducted over two days. Monologue Pair and Monologue Solo happened on Day 1, Dialogue Pair and Dialogue Solo on Day 2. As a result, one noise is that the Monologue

groups might share the information about the experiment with Dialogue groups before the Dialogue's turn, though they were not given answers and test scores. This could give Dialogue groups advantages, leading to larger learning gains than those of Monologue groups. However, this only happened between Dialogue Pair and Monologue Pair (d = 1.08, p = .004). The difference between Dialogue Solo and Monologue Solo was not significantly different (d = –0.33, p = .385). Furthermore, I collapsed Dialogue Pair and Dialogue Solo into one group Dialogue, the same with Monologue. The result of ANCOVA showed no significant difference in learning outcomes between Dialogue and Monologue, that is, between the two days, $F(1, 57) = 2.74$, p = .104, η_p^2 = .05. It indicates a low possibility of noises influencing experiment results.

One difference between the dialogue and the monologue videos is the length. The dialogue video is 25 minutes 16 seconds, longer than the monologue (18 minutes 22 seconds) due to interaction episodes. I would have to reduce the content in the dialogue to equalize the length of the two videos. I chose to control the content rather than the length because I presumed that the content had a greater influence on observers' learning. If the dialogue video is superior for its length, why did participants in Dialogue Solo have similar learning outcomes to those in Monologue Solo? There are not any plausible explanations. To control the effect of video length, future researchers could make two videos equivalent in length by adding irrelevant content (not included in tests) into a monologue.

Unlike some other studies (Chi et al. 2008; Craig et al., 2009; Muldner et al., 2014), I did not record participants' video watching and problem solving sessions due to limited resources. The information could help explore the factors that render Dialogue Pair more effective.

There was only a short-term measurement after the experiment. A second posttest several days or weeks after the experiment could offer more information about observers' learning

outcomes, that is, long-term retention. This decision was due to two reasons. First, I was not allowed to do another posttest because students' time in school was not available. Second, students might access the learning content from other sources before the second posttest, which would complicate the attribution of learning outcomes to experimental interventions.

Future Research

As discussed above, more studies are warranted to understand the interaction effect of dialogues and collaboration. Later experiments can inspect Chi et al.'s (2017) explanations by manipulating a tutee's behaviours and conflict episodes in dialogue videos. Also, Cooper et al.'s (2018) and my research shows that low- and high-achieving learners responded differently to observational learning conditions. Thus, there is a scope for future research on observers of different learning achievements. The present study differs from other experiments by removing the tutor's and the tutee's visual presence but came to the same findings. Future research may find a null effect of talk-heads in dialogue videos, expending the null effect of visual presence in monologue videos in Lee's (2019) experiment.

Also, my experiment did not have a human tutoring group as a benchmark because of limited resources. Both Chi et al. (2008) and Muldner et al. (2014) included a human tutoring condition and found that students who watched dialogue videos and collaborated in pairs performed comparably as tutees in the dialogue videos that they watched. As a result, Muldner et al. (2014) concluded that collaboratively observing dialogue videos had a higher utility than being tutored because, with similar effect sizes, human tutoring was a less scalable approach that demanded too many resources to be massively implemented. Future research can further examine the utility of tutorial dialogues by comparing Dialogue Pair with human tutoring.

As I know, this is the first study on tutorial dialogues outside English-speaking countries, drawing the same conclusions as many previous experiments. To my knowledge, there were only two prior studies on high school populations (Gholson et al., 2009; Muldner et al., 2014), while others were on college students. More studies on younger populations and in non-English-speaking cultures are needed to understand the generalizability of the effects of dialogues and collaboration on learners during observational learning.

So far, the disciplines of most studies on dialogue video are natural sciences, such as physics (Chi et al., 2008; Craig et al., 2009; Gholson et al., 2009; Muldner et al., 2014; Muller et al., 2007; Muller et al., 2008), computer science (Craig et al., 2000; Craig et al., 2004; Craig et al., 2006; Driscoll et al., 2003; Fox Tree, 1999; Gholson et al., 2009; Lee 2019), and physiology (Cooper et al., 2018). Regardless of different topics, those experiments are all consistent with my finding on the interaction effect between dialogue and collaboration. However, more investigations are required in other domains like hands-on skills, language, and arts before generalizing the interaction effect.

Though further research is required to understand the advantages of Dialogue Pair, it should not stop educators from harnessing the utility of this pedagogy. Currently, a monologue is a common format for tutorial videos. More dialogue videos can be created to allow collaborative learners to optimize their learning. However, when implementing this method, educators need to consider factors that impact its effectiveness, such as the difficulty of video content for observers, tutees' active behaviours, and tutors' pedagogy like scaffolding in the videos. We may discover more about tutorial dialogues not only from empirical research but also from classroom practice.

Conclusion

To conclude, the present study replicates most previous findings. It shows that compared with watching the dialogue video solo, or watching the monologue video solo or in pairs, collaborative observation of the dialogue video is the most effective approach for learning a difficult psychology topic among a junior high Chinese population. My results reveal a strong interaction effect of dialogues and collaboration on learning. Modelling and motivation explain the results but have to be tested in future experiments. The questionnaire responses show that observers and the tutee had close perspectives and shared misconceptions about the learning content. There is a scope for further investigations on tutorial dialogues with large samples in other disciplines and different populations.

References

Almasri, F., Hewapathirana G.I., Ghaddar, F., Lee, N., & Ibrahim, B. (2021). Measuring attitudes towards biology major and non-major: Effect of students' gender, group composition, and learning environment. *PLOS ONE 16*(5): e0251453. https://doi.org/10.1371/journal.pone.0251453

Andrist, L., Chepp, V., Dean, P., & Miller, M. V. (2014). Toward a video pedagogy: A teaching typology with learning goals. *Teaching Sociology, 42*(3), 196–206.

Bailey, E. G., Greenall, R. F., Baek, D. M., Morris, C., Nelson, N., Quirante, T. M., Rice, N. S., Rose, S., & Williams, K. R. (2020). Female in-class participation and performance increase with more female peers and/or a female instructor in life sciences courses. *CBE—Life Sciences Education 19*(3), ar30. https://doi.org/10.1187/cbe.19-12-0266

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology, 1*(6), 589–595. https://doi.org/10.1037/h0022070

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology, 63*(3), 575–582. https://doi.org/10.1037/h0045925

Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. *The Journal of Abnormal and Social Psychology, 66*(1), 3–11. https://doi.org/10.1037/h0048687

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

Chi, M. T. H. (2013). Learning from observing an expert's demonstration, explanations, and dialogues. In J. J. Staszewski (Ed.), *Expertise and skill acquisition: The impact of William G. Chase* (1st ed., pp.1–27). Taylor & Francis Group. https://doi.org/10.4324/9780203074541

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences, 26:1*, 10–50. https://doi.org/10.1080/10508406.2016.1204546

Chi, M. T. H., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, *32*(2), 301–341. https://doi.org/10.1080/03640210701863396

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1

Cooper, K. M., Ding, L., Stephens, M. D., Chi, M. T. H., & Brownell, S. E. (2018). A course-embedded comparison of instructor-generated videos of either an instructor alone or an instructor and a student. *CBE—Life Sciences Education*, *17*(2), ar31. https://doi.org/10.1187/cbe.17-12-0288

Cox, R., McKendree, J., Tobin, R., Lee, J., & Mayes, T. (1999). Vicarious learning from dialogue and discourse. *Instructional Science, 27*(6), 431–458. https://doi.org/10.1023/A:1003489631631

Craig, S. D., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*, *101*(4), 779–789. https://doi.org/10.1037/a0016601

Craig, S. D., Driscoll, D. M., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia*, *13*(2), 163–183. https://www.learntechlib.org/primary/p/24271/

Craig, S. D., Gholson, B., Ventura, M., Graesser, A. C., & the Tutoring Research Group. (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *International Journal of Artificial Intelligence in Education, 11*, 242–253. https://people.cs.pitt.edu/~litman/courses/slate/pdf/paper.pdf

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, *24*(4), 565–591. https://doi.org/10.1207/s1532690xci2404_4

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20,* 159–165. https://content.iospress.com/articles/work/wor00285

Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, *29*(4), 431–450. https://doi.org/10.2190/Q8CM-FH7L-6HJU-DT9W

Dugard, P., & Todman, J. (1995). Analysis of pre-test-post-test control group designs in educational research. *Educational Psychology, 15*(2), 181–198. https://doi.org/10.1080/0144341950150207

Fox Tree, J. E. (1999). Listening in monologs and dialogues. *Discourse Processes, 27*(1), 35–53. https://doi.org/10.1080/01638539909545049

Geertshuis, S., Rix, N., Murdoch, O., & Liu, Q. (2021). Learning by watching others learn: Vicarious learning from videoed tutorials. In D. S. P. Gedera & A. Zalipour (Eds), *Video pedagogy: theory and practice* (1st ed., pp. 103–122). Springer, Singapore. https://doi.org/10.1007/978-981-33-4009-1

Gholson, B., Witherspoon, A., Morgan, B., Brittingham, J. K., Coles, R., Graesser, A. C., Sullins, J., & Craig, S. D. (2009). Exploring the deep-level reasoning questions effect during vicarious learning among eighth to eleventh graders in the domains of computer

literacy and Newtonian physics. *Instructional Science*, *37*(5), 487–493. https://doi.org/10.1007/s11251-008-9069-2

Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist, 45*(4), 234–244. https://doi.org/10.1080/00461520.2010.515933

Harskamp, E., Ding, N., & Suhre, C. (2008). Group composition and its effect on female and male problem-solving in science education. *Educational Research, 50*(4), 307–318. https://doi.org/10.1080/00131880802499688

JASP Team. (2021). *JASP (Version 0.16)* [Computer software]. JASP. https://jasp-stats.org/

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46–53. https://doi.org/10.3102/0013189X15569530

Lee, B. J. (2019). *An eye tracking comparison of instructional videos showing a monologue versus a dialogue: Impacts on visual attention, learning, and psychological variables* [Master's thesis, Carleton University]. Carleton University Research Virtual Environment. https://doi.org/10.22215/etd/2019-13787

Lee, B., & Muldner, K. (2020). Instructional video design: Investigating the impact of monologue- and dialogue-style presentations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3313831.3376845

Muldner, K., Lam, R., & Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, *106*(1), 69–85. https://doi.org/10.1037/a0034448

Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning*, *24*(2), 144–155. https://doi.org/10.1111/j.1365-2729.2007.00248.x

Muller, D. A., Sharma, M. D., Eklund, J., & Reimann, P. (2007). Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science, 35*(6), 519–533. https://doi.org/10.1007/s11251-007-9017-6

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. https://doi.org/10.1016/0010-0285(89)90008-X

Shenzhen Tencent Computer System Co., Ltd. (2022). *QQ Windows (Version 9.5.6)* [Computer software]. Tencent. https://im.qq.com/download

The G*Power Team. (2020). *G*Power (Version 3.1.9.7).* [Computer software]. Heinrich Heine University Düsseldorf. https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Zhan, Z., Fong, P. S., Mei, H., & Liang, T. (2015). Effects of gender grouping on students' group performance, individual achievements and attitudes in computer-supported collaborative

learning. *Computers in Human Behavior 48*, 587–596.

https://doi.org/10.1016/j.chb.2015.02.038

Appendices

(Appendix 1-4 were translated from Chinese.)

Appendix 1

Pre- and Post Test

1. Which of the following options are correct?
Eating fatty fish helps to absorb calcium
Eating fatty fish can supplement Omega 3 fatty acids
Eating fish with less fat is better for memory
Fatty fish usually exist in freshwater

2. Regarding vitamin D, which of the following statements is correct?
Sun exposure can promote the absorption of vitamin D by the body
We can only get vitamin D from food
Vitamin D helps maintain the stability of the neural network in the hippocampus
Vitamin D helps the production of new nerve cells

3. Which of the following statements is correct?
We can get a lot of vitamin D from walnuts
We can get a lot of vitamin D from soy milk
We can get a lot of Omega 3 fatty acids from milk
We can get a lot of Omega 3 fatty acids from egg yolk

4. During deep sleep period,
People are more likely to dream, and dreams are clearer
It is an important stage of declarative memory consolidation
The muscles will be temporarily paralyzed
Muscle occasionally twitches

5. You learned to play tennis. During which sleep stage, if you are disturbed, your memory of playing tennis will likely get worse?
Deep sleep
REM period
Stage 1
Stage 2

6. Regarding the hippocampus, which of the following is correct
The hippocampus is the place for memory processing
Short-term memory information is transferred from the cerebral cortex to the hippocampus and stored in the form of long-term memory

The hippocampus is essential for procedural memory information processing
There are neural networks in the hippocampus
With frequent practice of the same movement, the hippocampus will regrow

7. Which of the following statements is correct
60% of the human brain is fat and fatty acids
Declarative memory is about “what it is”
Procedural memory is about “how to do it”
The hippocampus will swell if someone often drinks much alcohol

8. Please select the wrong option:
When our emotional experience is strong, the information we receive is easier to remember
When our emotional experience is strong, the received information is transmitted to the hippocampus faster
Regular review can help us convert short-term memory into long-term memory and store it in the hippocampus
Regular review can help form more short-term memory

9. What might happen if vitamin D is lacking?
You often want to sleep
Skin will become pale
Memory will get poor
The neural network in the hippocampus is easily eroded by viruses
The neural network in the hippocampus is easily destroyed by biological enzymes

10. What happens when a person gets drunk frequently?
Poor memory
Less vitamin D in the body
The neural network in the hippocampus is destroyed
The hippocampus becomes smaller
The production of new nerve cells in the hippocampus is disturbed

11. What happened after the patient’s hippocampus was removed?
He has no short-term memory
He has no long-term memory
He can’t remember newly learned physical movements
He cannot remember newly learned abstract concepts

12. Which of the following activities mainly rely on declarative memory?
Know which month it is now
Learn the content in the video
Learn to swim
Practice calligraphy
Play the piano
Recall the first biking experience
Play basketball

13. Regarding sleep and memory, which of the following statements is correct
Consolidation of declarative memory occurs during deep sleep
Consolidation of procedural memory occurs during deep sleep
Consolidation of declarative memory occurs during REM phrase
Consolidation of procedural memory occurs during the REM phase

14. The following statement is wrong
The information we receive from the outside world is first stored in the hippocampus
Frequent review will increase our memory capacity, so that we can remember longer
When we do not review something and have a weak emotional experience of it, the memory will gradually disappear in the cerebral cortex
When we often recall something, it's hard to forget it

15. The following statement is wrong
Vitamin D promotes the conversion of short-term memory into long-term memory
Vitamin D will help the transmission of memory information
Vitamin D can protect the biological enzymes in the hippocampus
Vitamin D can maintain the stability of neural networks in the hippocampus

Appendix 2

Questionnaire

1. Was there a student's voice in the video that you watched?
○Yes
○No (jump to Question 8)

2. I like this video.
○😄 ○🙂 ○😐 ○😕 ○😣

3. Why? Please tell us the reason for your choice:
_____

4. The student in the video was helpful to my watching and learning.
○😄 ○🙂 ○😐 ○😕 ○😣

5. The student in the video interfered with my watching and learning.
○😄 ○🙂 ○😐 ○😕 ○😣

6. I hope that there is no student' voice in the video.
○😄 ○🙂 ○😐 ○😕 ○😣

7. Please explain your choice for Question 4, 5, and 6:

_____

8. I like this video.
○😁 ○🙂 ○😐 ○😕 ○😣

9. Why?

_____

10. The teacher in the video was helpful to my watching and learning.
○😁 ○🙂 ○😐 ○😕 ○😣

11. I was very focused while watching the video.
○😁 ○🙂 ○😐 ○😕 ○😣

12. Why?

_____

13. Rather than watching a video, I hope to interact with the teacher directly.
○😁 ○🙂 ○😐 ○😕 ○😣

14. Why?

_____

15. Did you just watch a video and discuss the Learning Task with a partner?
○Yes
○No (jump to Question 18)

16. I like to watch the video and discuss problems with my partner.
○😁 ○🙂 ○😐 ○😕 ○😣

17. Why?

_____

18. Please share your thoughts about the video and your participation:

_____

Appendix 3

Learning Task

1. How is memory formed?
2. How do Omega 3, vitamin D, and alcohol affect memory?

3. Please list six activities you usually do. Which involves declarative memory, and which involves procedural memory? In what stages are these memories consolidated?
4. What can you do to help memorize things?
5. What did you learn from this video?

Appendix 4

Dialogue and Monologue Excerpts

Dialogue:
Tutor: …He couldn't remember what he just read in books, but he could still remember some body movements after repetitive practice. What does this case mean?
Tutee: It means with frequent practice of the same movement, the hippocampus will regrow.
Tutor: But the hippocampus was removed, and it couldn't grow again.
Tutee: It was the result of repetition.
Tutor: Actually, this case indicates that there are two types of memory, declarative memory and procedural memory.

Monologue:
Tutor: …He couldn't remember what he just read in books, but he could still remember some body movements after repetitive practice. What does this case mean? Does this mean with frequent practice of the same movement, the hippocampus will regrow? Actually not. The hippocampus cannot regrow after being removed. This case indicates that there are two types of memory, declarative memory and procedural memory.

Appendix 5

The Number of Participants Making Mistakes on Misconception 2, 3, 4, 5 in the Posttest After Watching the Dialogue/Monologue Videos in Pairs/Solo

Mis 2	Right	Wrong	$\chi^2(1)$	*p*
Mono	9	20	0.36	0.550
Dia	8	25		
Mono Pair	8	10	2.62	0.106
Dia Pair	4	16		
Mono Solo	1	10	1.70	0.193
Dia Solo	4	9		
Pair	12	26	0.85	0.356
Solo	5	19		
Dia Pair	4	16	0.50	0.481
Dia Solo	4	9		
Mono Pair	8	10	3.99	0.046*
Mono Solo	1	10		

Mis 3	Right	Wrong	$\chi^2(1)$	*p*
Mono	17	12	1.62	0.203
Dia	14	19		
Mono Pair	11	7	1.69	0.194
Dia Pair	8	12		
Mono Solo	6	5	0.17	0.682
Dia Solo	6	7		
Pair	19	19	0.00	1.000
Solo	12	12		
Dia Pair	8	12	0.12	0.727
Dia Solo	6	7		
Mono Pair	11	7	0.12	0.728
Mono Solo	6	5		

Mis 4	Right	Wrong	$\chi^2(1)$	*p*
Mono	19	10	0.16	0.690
Dia	20	13		
Mono Pair	12	6	0.32	0.572
Dia Pair	15	5		
Mono Solo	7	4	1.51	0.219
Dia Solo	5	8		
Pair	27	11	2.79	0.095
Solo	12	12		
Dia Pair	15	5	4.41	0.036*
Dia Solo	5	8		
Mono Pair	12	6	0.03	0.868
Mono Solo	7	4		

Mis 5	Right	Wrong	$\chi^2(1)$	*p*
Mono	17	12	0.83	0.363
Dia	23	10		
Mono Pair	12	6	0.32	0.572
Dia Pair	15	5		
Mono Solo	5	6	0.62	0.431
Dia Solo	8	5		
Pair	27	11	1.83	0.176
Solo	13	11		
Dia Pair	15	5	0.68	0.411
Dia Solo	8	5		
Mono Pair	12	6	1.27	0.260
Mono Solo	5	6		

Note. * indicates significant p-values ($p < 0.05$).

Appendix 6

Summary of Relevant Studies

Authors	Domain	Sample	Finding
Chi et al., 2008	quantitative kinematics	70 undergraduates	tutoring = dialogue in pairs > dialogue alone / collaborating / study alone
Cooper et al., 2018	physiology	280 university students	Low GPA students: monologue > dialogue; Others: monologue = dialogue
Cox et al., 1999	sentence parsing	54 undergraduates	dialogue = monologue
Craig et al., 2000	computer literacy	48 university students	dialogue > monologue (Viewers asked more questions [sign of deep reasoning]. No learning outcome was measured.)
Craig et al., 2004	computer literacy	university students: 120 (Exp 1), 110 (Exp 2)	tutoring > dialogue in pairs (few conversations) > dialogue alone (not statistically significant)
Craig et al., 2006	computer literacy	140 undergraduates (Exp 2)	deep questions dialogue = deep questions monologue > tutoring = dialogue
Craig et al., 2009	rotational kinematics	67 college students	dialogue in pairs > monologue in pairs = monologue alone (long-term measures)
Driscoll et al., 2003	computer literacy	university students: 48 (Exp 1), 96 (Exp 2)	Exp 1: overhearing dialogue (with deep questions) > monologue; Exp 2: dialogue with deep questions > monologue/ dialogue with the virtual tutee's shallow questions or comments
Fox Tree, 1999	arrange tangram figures	167 university students	overhearing dialogue > monologue
Gholson et al., 2009	computer literacy, Newtonian physics	8,9,10,11 graders	dialogue (deep questions) > monologue (without) = virtual tutoring
Lee, 2019	Python	77 undergraduates	dialogue = monologue

Muldner et al., 2014	diffusion	50 undergraduates (Exp 1), 40 junior high (Exp 2)	Exp 1: tutoring = dialogue in pairs > monologue in pairs; Exp 2: tutoring > dialogue in pairs = monologue in pairs
Muller et al., 2007	quantum mechanics	second-year physics class	dialogue with misconceptions > monologue without
Muller et al., 2008	Newtonian first and second laws of motion	678 first-year undergraduates	dialogue with misconceptions = monologue with misconceptions > monologue without misconceptions
Schober &. Clark, 1989	arrange tangram figures	20 + 40 Stanford university students	tutoring > overhearing dialogue

Appendix 7

Ethics Approval

MEMORIAL UNIVERSITY

Interdisciplinary Committee on Ethics in Human Research (ICEHR)

St. John's, NL Canada A1C 5S7
Tel: 709 864-2561 icehr@mun.ca
www.mun.ca/research/ethics/humans/icehr

ICEHR Number:	**20221813-ED**
Approval Period:	November 18, 2021 – November 30, 2022
Funding Source:	
Responsible Faculty:	Dr. Timothy Seifert Faculty of Education
Title of Project:	*Learning by Watching Tutorial Videos: The Impact of Dialogue and Collaboration*

November 18, 2021

Mr. Xingbang Chen
Faculty of Education
Memorial University of Newfoundland

Dear Mr. Chen:

Thank you for your correspondence addressing the issues raised by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) for the above-named research project. ICEHR has re-examined the proposal with the clarifications and revisions submitted, and is satisfied that the concerns raised by the Committee have been adequately addressed. In accordance with the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2)*, the project has been granted *full ethics clearance* for **one year**. ICEHR approval applies to the ethical acceptability of the research, as per Article 6.3 of the *TCPS2*. Researchers are responsible for adherence to any other relevant University policies and/or funded or non-funded agreements that may be associated with the project. If funding is obtained subsequent to ethics approval, you must submit a Funding and/or Partner Change Request to ICEHR so that this ethics clearance can be linked to your award.

The *TCPS2* **requires** that you **strictly adhere to the protocol and documents as last reviewed** by ICEHR. If you need to make additions and/or modifications, you must submit an Amendment Request with a description of these changes, for the Committee's review of potential ethical concerns, before they may be implemented. Submit a Personnel Change Form to add or remove project team members and/or research staff. Also, to inform ICEHR of any unanticipated occurrences, an Adverse Event Report must be submitted with an indication of how the unexpected event may affect the continuation of the project.

The *TCPS2* **requires** that you submit an Annual Update to ICEHR before **November 30, 2022**. If you plan to continue the project, you need to request renewal of your ethics clearance and include a brief summary on the progress of your research. When the project no longer involves contact with human participants, is completed and/or terminated, you are required to provide an annual update with a brief final summary and your file will be closed. All post-approval ICEHR event forms noted above must be submitted by selecting the ***Applications: Post-Review*** link on your Researcher Portal homepage. We wish you success with your research.

Yours sincerely,

James R. Drover

James Drover, Ph.D.
Vice-Chair, Interdisciplinary Committee on
Ethics in Human Research

JD/bc

cc: Supervisor – Dr. Timothy Seifert, Faculty of Education