Online Collaborative Video Viewing (CVV):

The Impact of Collaborative Modes in Active Video-based Learning

By

Ratrapee Techawitthayachinda

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2023 by the
Graduate Supervisory Committee:

Michelene Chi, Co-Chair
Yi-Chun Hong, Co-Chair
Brian Nelson

ARIZONA STATE UNIVERSITY

December 2023

ABSTRACT

Collaborative Video Viewing (CVV) transforms passive video-based learning into an engaging, active process. While collaborative modes have different affordances that could potentially influence knowledge co-construction, no study has directly assessed the impact of collaborative modes in CVV activities. Therefore, this current study seeks to investigate how collaborative modes influence learning outcomes, learning engagement, group interaction and the co-construction process.

The study utilized a within-subject, counterbalanced experimental design, in which each participating undergraduate student was paired in dyads. These dyads were assigned to engage in two separate CVV sessions: one using synchronous voice-based collaborative mode (SV) and the other using asynchronous text-based collaborative mode (AT). After each session, participants completed a test consisting of retention and application questions. ANCOVA was utilized to analyze the test scores.

To ascertain if the different scores were a result of varying levels of learning engagement, dyad discussions were coded using ICAP coding (Chi & Wylie, 2014). Furthermore, to delve deeper into the group interaction mechanism in SV and AT, a codebook was developed to analyze the discourse that occurred during dyad interaction. Sequential analysis and thematic narrative analysis were employed to visualize interaction patterns and the co-construction process.

The findings indicated that, generally, SV dyads performed better on application scores and have significantly higher interactive learning engagement than AT dyads. In line with ICAP predictions, the higher-score groups in both SV and AT engaged in more generative processes, leading to more constructive and interactive comments than lower-scoring groups. In terms of group interaction, both SV and AT primarily use descriptive discourse for co-explanation. However, the SV groups exclusively introduce discourse expressing uncertainty, which subsequently leads to group negotiation. The study identified distinct knowledge co-construction phases, including (a) co-explanation, (b) negotiation, and (c) application. Although the co-explanation phase is the most frequent in all dyad scores in both SV and AT, the negotiation phase appears to differentiate low-high score dyads from high-high score dyads.

These findings hold research implications for understanding learning engagement and group interaction in various online collaborative modes, as well as for the instructional design of active video-based learning through collaborative video viewing.

DEDICATION

I dedicate this dissertation to my loving family for their unwavering support and encouragement throughout my academic journey. To my professors and mentors who have inspired and guided me, your wisdom has been invaluable. This work is a tribute to all of you!

ACKNOWLEDGMENTS

My Ph.D. journey has been shaped by the unwavering support and inspiration of some exceptional individuals.

First and foremost, I'm deeply grateful to Dr. Hong. From day one, your wisdom and encouragement have been a constant source of strength throughout this challenging yet rewarding journey. I extend my heartfelt thanks to Dr. Chi for being a continual source of influence beyond the impactful ICAP framework. Your character, intelligence, and work ethic have consistently inspired my growth as a researcher. Dr. Nelson's contributions to modality literature not only significantly influenced the foundation of my dissertation but also exemplified his kindness and unwavering support.

I'm also grateful to the professors who played vital roles in my Ph.D. journey. Dr. Chen's guidance in uncertainty management enriched my understanding of group interactions and qualitative research. Dr. Benedict's inclusive research group offered valuable K-12 research experience. I appreciate all the professors who taught me in various courses and collaborated on research. I'd like to give special thanks to Ian and Amanda, my supervisors at the Center for the Future of Arizona, for their professional, respectful, and supportive environment during the latter years of my Ph.D. journey.

Lastly, my deepest gratitude goes to my parents, Dr. Wirat and Suriyaporn Techawitthayachinda, my husband Dr. Rafael Iriya, and our beloved babies BaoBao and PanPan. They have all played a monumental role in shaping who I am today, providing love, support, and understanding that have fortified my current self.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Definition of Terms

This writing employs specialized vocabulary throughout, some of which may not be well-known to all readers. As a result, the following definitions for frequently recurring terms are provided below before commencing the narrative.

Collaborative Video Viewing (CVV): An instructional activity in which students watch the same video and discuss related content to accomplish shared tasks (e.g., explaining concepts or solving problems).

Asynchronous Text-based Collaborative Mode (AT): A collaborative environment that involves flexible time, text-based communication.

Synchronous Voice-based Collaborative Mode (SV): A collaborative environment that involves real-time, voice-based communication.

Social Video Annotation Tool: An asynchronous text-based collaborative tool that allows participants to bookmark or highlight specific portions of video content to add comments or initiate discussions.

Video Conferencing Tool: A synchronous voice-based collaborative tool that enables participants to communicate through video in real-time.

Dyad-Score Groups: Dyads that are categorized into groups based on performance of both partners.

Learning Engagement: Following the ICAP framework, the quality of learning depends on how a pair of students overtly engage with learning materials and with each other.

Discourse: The language and communication students employ in dialogue to construct knowledge (e.g., asking questions, seeking clarification, etc.).

Discourse Patterns: The frequency of discourse sequences used in dialogue to construct knowledge (e.g., clarification → elaboration).

Overview

Online Video-based Learning

Improvement of technology and the increased demand for flexible learning makes online learning grow and change the current educational system. Online learning is no longer limited to distance learning programs in higher education but becomes a part of the on-campus programs and traditional classrooms (e.g., blended learning and flip classroom) (Baker, 2016; Seaman et al., 2018). Recently, online learning has moved from an option to a necessity due to the outbreak of the COVID pandemic, requiring the inevitable shift from face-to-face to the online classroom (Dhawan, 2020).

Video-based learning is a primary instructional method commonly employed in online learning. For example, Massive Open Online Courses (MOOCs) incorporate high-quality videos from various universities to cater to students worldwide (Deng & Gao, 2023; Hansch et al., 2015; Hew, 2016). More recently, instructors have explored the possibility of reusing existing recorded lectures during urgent transitions to online learning prompted by the COVID-19 outbreak (Dhawan, 2020). Lecture videos allow educational institutions to provide high-quality instruction on demand, enhance the learning experience, and offer scalability and cost-effectiveness. Simultaneously, video recordings provide students with a wide range of autonomy and flexibility regarding time, place, and cognitive resource allocation (Costley et al., 2021; Wong et al., 2022; Schwan & Riempp, 2004). From a pedagogical standpoint, video-based learning is particularly effective in visual strategies (seeing), vicarious experiences (doing), and motivation (engaging) (de Koning et al., 2018; Koumi, 2006).

However, major critiques of online video-based learning evolve around "recording absorption". Lecture videos predominantly follow a talking-head, lecture-style format and are limited to an individual-level task. Students spend a significant amount of time intensely watching videos in solitude, experiencing isolation and feeling confined to the screen, passively participating in information absorption rather than knowledge construction (Chi et al., 2017; Hansch et al., 2015; Zhang et al., 2020). In reality, mind-wandering is prevalent during video lectures. Most students lose concentration for approximately 40% of the time during a video lasting 20 minutes or longer (Kane et al., 2017; Szpunar et al., 2013; Zhang et al., 2020). These passive behaviors are linked to low learning engagement and suboptimal learning outcomes (Chi & Wylie, 2014, Yoon et al., 2021). Recording high-quality lectures may achieve *efficiency*, executing existing processes in a more sustainable or scalable manner). However, without aligning technology and pedagogy, the *enhancement* or *transformation* of learning, improving or altering existing processes and outcomes, is challenging to achieve (Kirkwood & Price, 2014; Suthers, 2006).

Overcome the Challenges

Learning engagement and outcomes in video-based learning can be leveraged through additional learning strategies and instructional design support. Multiple factors have been examined to overcome the challenges to reduce cognitive load, maximize student engagement, and promote active learning in video-based learning (Brame, 2016; Fiorella & Mayer, 2016, Navarrete et al., 2021; Sablić et al., 2021). Strategies may target (a) video production such as the instructional design on video formats (e.g., Chi et al., 2017; Crook & Schofield; 2017; Hansch et al., 2015) and multimedia design on videos (e.g., Mayer, 2021); (b) user-interface interaction such as a level of learner's control on video instruction (e.g., Schroeder et al, 2020) and interface design for scaffolding and other interactive features (e.g., Mamun et al., 2020; Seo et al., 2021); and (c) instructional activity and sequences around video-based instruction (e.g., Muldner et al., 2011; Fiorella et al., 2020). *Collaborative video viewing (CVV)* is one of the instructional activities designed to use social facilitation to engage students in active learning. It involves having students watch instructional videos and engage in discussions to collectively achieve shared learning objectives.

Back in the 70s, the concept of social watching in education originated with *Tutored Video Instruction (TVI).* Gibbons and colleagues (1977) introduced TVI, an activity in which remote students locally watch lecture videos and discuss them with a peer and a tutor. This social-watching video activity not only provides remote students access to high-quality video instruction but also mitigates the negative effects of individual lecture-video viewings, such as passive watching and mind-wandering. Even without tutors, other studies subsequently demonstrated the advantages of collaborative video viewing (CVV) of lectures among students themselves (e.g., Li et al., 2014; Liu et al., 2023; Pi et al, 2022; Weisz et al., 2007). Thus, the significant benefits of social watching stem not only from tutoring itself but also from social interaction and the opportunity for constructive engagement during discussions (Chi et al., 2017; Pi et al, 2022;). Understanding is fostered through participation in dialogues to establish common ground (Schober & Clark, 1989; Stahl, 2006), as evidenced by the sustainability of learning effectiveness in peer-to-peer groups.

Previous studies conducted in face-to-face classrooms and laboratories have consistently shown that Collaborative Video Viewing (CVV) leads to a superior learning experience and better outcomes compared to solitary watching and various other individual learning strategies. (e.g., Chi et al, 2008; Li et al., 2014; Muldner et al., 2011; Weisz et al., 2007). For instance, Chi et al. (2008) conducted a study comparing CVV on tutorial dialog videos with other instructional strategies such as tutoring, collaborative problem-solving after reading, viewing video alone, and studying alone. Their findings revealed that collaborative activities involving student interaction (e.g., tutoring, CVV, collaborative problem-solving) yielded superior learning outcomes compared to individual activities (e.g., viewing video alone and studying alone).

As both the CVV and alone-viewing groups watched the same tutorial dialog videos, the CVV group's superior performance in physics problem-solving highlights the collaborative effect in the co-watching context. Chi et al. (2008) further expounded on the comparable outcomes of CVV on tutorial dialog videos and tutoring, stating that "(collaborative) observing can be as effective a way to learn as tutoring if participants have opportunities to be active and constructive." They also emphasized the scalable potential of CVV, indicating that "one can be active and constructive by interacting with a peer and not necessarily with a tutor" (p. 315). The discussion surrounding videos shifts them from passive learning resources to generative stimuli for knowledge construction in active learning pedagogy, promoting externalization, perspective-taking, reflection, and critical thinking (Reisman & Enumah, 2020; Sherer & Shea, 2011; Agarwala et al., 2012).

Statement of the Problem

Over the decades, substantial research has demonstrated the effectiveness of student-tutor and peer-to-peer collaborative video viewing in face-to-face contexts. However, limited research has been conducted on peer-to-peer collaborative video viewing (CVV) to examine its effectiveness on learning outcomes in an online setting. The replicability of learning results while transitioning pedagogically from one environment to another is not always certain, and the similarity of learning experiences and outcomes between online and face-to-face learning environments remains inconclusive (Greenhow et al., 2022; Henriksen et al., 2020; Lyons, Reysen, & Pierce, 2012). The implementation of learning activities in an online setting, including the selection of communication tools and their potential impact on the learning process and outcomes, remains underexplored and lacks comprehensive understanding (Greenhow et al., 2022).

Emerging from conventional threaded discussions and text annotation, social annotation has become a prevalent asynchronous tool in educational Collaborative Video Viewing (CVV) endeavors. In social annotation, students can highlight specific segments of a video and add annotations, with these highlights visible to the entire group. Annotations can take the form of structured notes (e.g., Fang et al., 2022), or engage in dialogues and conversational styles (e.g., Agarwala et al., 2012; Lam & Habil, 2021). Moreover, social engagement levels can encompass a spectrum, extending from the simple display of annotations to make group cognition transparent to discussions embedded within annotations to foster focused discussions around videos.

Several educational studies have investigated the instructional designs of social annotation tools for online CVV, with a particular emphasis on their application in behavioral modeling and reflection. This includes fostering students' reflective practices and critical thinking while engaging with various types of videos, such as teacher teaching, sport training, and medical training (Evi-Colombo et al., 2023; Rich & Hannafin, 2009; Vohle & Reinmann, 2014). Despite meta-analysis affirming that instructional videos not only enhance practical, observable skill learning but also facilitate cognitive skill and knowledge acquisition (Lin & Yu, 2023), there has been limited research conducted to explore the utilization of collaborative tools in CVV for these areas.

With the improvement of internet speed and accessibility, synchronous voice-based collaborative tools, such as video conferencing, play an increasingly vital role in education. However, video conferencing tools primarily serve as a primary channel for educational purposes, encompassing activities such as meetings, team roleplay, and discussions (Geets et al., 2011; Singh & Thurman, 2019). Only a limited number of studies utilize video conference tools for a backchannel mode of communication while simultaneously engaging in core activities like instructional video co-watching and educational games (e.g., Cadiz et al., 2000; Erlandson et al., 2010; Liu et al., 2023). In the realm of human-computer interaction (HCI), a substantial body of literature investigates the usability of application configurations and perceptions of synchronous co-watching tools on shared entertainment experiences (Geerts et al., 2011). Although studies show synchronous co-watching tools enhance watching experience, the learning implications and outcomes of integrating videoconferencing into online educational Collaborative Video Viewing (CVV) remain both underexamined and insufficiently informed (Cadiz et al., 2000; Habes et al., 2020).

In summary, this introductory chapter has identified a critical gap in existing literature: the lack of a direct comparison between asynchronous text-based collaboration (social annotation) and synchronous voice-based collaboration (videoconferencing) in the context of online Collaborative Video Viewing (CVV) for educational purposes. This study aims to address this gap by examining how these collaborative modes and tools impact the learning process and outcomes in CVV, with the following research questions.

Research Questions

1. Are there differences in learning outcomes, between individuals and dyad-score groups in synchronous voice-based and asynchronous text-based collaborative watching?
2. What are the differences in learning, between individuals and dyad-score groups in voice-based synchronous and asynchronous text-based collaborative watching?

3. What kind of discourse and discourse patterns do dyad-score groups employ to co-construct knowledge in voice-based synchronous and asynchronous text-based collaborative watching?

In the upcoming chapters, we will delve into our research methodology, data collection and analysis procedures, findings, and the implications arising from our investigation. This study seeks to offer insights that can enhance the effectiveness of online collaborative learning environments and provide evidence-based guidance to educators and institutions for optimizing the use of collaborative modes and tools in online CVV.

CHAPTER 2

BACKGROUND LITERATURE

This chapter is organized into two distinct sections: the theoretical framework and the literature review. In the first section, the theoretical framework lays the foundation for the research by presenting the theoretical concepts and models that underpin the study. Following that, the literature review section critically examines existing research and scholarly works related to the research topic, providing a comprehensive overview of the current state of knowledge in the field.

Theoretical Framework

"We look, but we don't really see; we hear, but we don't really listen"

2.1 ICAP Framework to Enhance Video-Based Learning

Video-based learning is much more than watching high-quality videos. Most of the research in video-based learning aims in the same direction: shifting students from passively watching videos to actively engaging with the video content. This endeavor aligns with active learning, which reinforces educational activities that go beyond merely attending traditional lectures passively, to increase engagement, enhance retention and understanding, and foster higher-order thinking (Fiorella et al., 2020; Freeman et al., 2014; Sablić et al., 2021).

To operationalize active learning, Chi (2009a) introduced the ICAP framework to categorize active learning across different levels of overt behaviors. The ICAP framework asserts that, alongside Passive learning (P), active learning can be further categorized into Active (A), Constructive (C), and Interactive (I). The ICAP framework operationalizes the idea that the quality of learning depends on overt activities and their corresponding cognitive engagement levels, following the order: Passive (P) < Active (A) < Constructive (C) ≤ Interactive (I) (Chi & Wylie, 2014). Learning activities positioned higher in this hierarchy theoretically lead to more sophisticated cognitive engagement, resulting in enhanced learning outcomes. The ICAP explains how learning engagement is associated with overt behaviors and why they result in different learning outcomes. The framework has been used in various contexts to bridge the gap between the learning process and outcomes, as well as theory and practice.

On one hand, the ICAP can be employed to design learning activities and predict learning outcomes based on the activities provided within video-based learning. For instance, as previously mentioned, students often passively watch videos and might even experience wandering thoughts if video-watching constitutes their sole learning activity.

However, instructors can eliminate this passive behavior by incorporating active elements into video-based instruction. For instance, students can engage at the active level (A) if they have control over video content (e.g., interactive video) (e.g., Schroeder & Craig, 2017; Schwan & Riempp, 2004) or if they are instructed to take notes or summarize video content (e.g., Delen et al., 2014). Constructive learning strategies (C), including prompts for explanation, reflection, and problem-solving, encourage students to go beyond provided materials and formulate generalizations (Fiorella & Mayer, 2016). Collaborative work, demanding contributions from all participants, offers an opportunity for co-construction or interactive engagement (I).

Unlike the self-induced nature of constructive engagement (C), interactive engagement is driven by external contributions, such as knowledge building, co-construction, and co-explanation. Students derive learning benefits from co-inference as they infer new knowledge from activated and integrated information and iteratively draw knowledge from peer interactions (Chi & Wylie, 2014; Deiglmayr & Spada, 2010). The interactive endeavor becomes an ideal core of collaborative learning, where students leverage social interaction to enhance their knowledge. Theoretically, interactive activities yield equal or greater learning outcomes compared to constructive activities, yet both types of learning activities surpass the others.

On the other hand, the ICAP framework serves as an analytical tool to gauge engagement levels demonstrated by students through overt behaviors. For instance, while a pair of students may collaborate on tasks that ideally fall under Interactive activity, they might exhibit different behaviors and levels of engagement (See Chi & Menekse, 2015). Beyond its predictive capacity, this study also utilized the ICAP to measure learning engagement occurring within group interactions, explaining why students attending the same collaborative learning activities might yield diverse learning results.

2.2 Cognitive Constructivism and Social Constructivism

In order to gain a comprehensive understanding of the advantages tied to interactive engagement within collaborative instructional design, it's crucial to revisit the social constructivism embedded within the broader socio-cultural context and the interplay between language or discourse, social interaction, and cognition.

Social constructivism, a facet of the broader socio-cultural theory, underscores the co-construction of knowledge through dialogue, collaborative problem-solving, and shared experiences. Social constructivism emphasizes that knowledge is collaboratively built through social interactions. These interactions involve cognitive processes of externalizing and internalizing information, which are contextualized through language and communication (Cress & Kimmerle, 2008; Jorczak, 2011; Kalina & Powell, 2009; Stahl, 2006). Both cognitive and social constructivism are rooted in the idea of constructing knowledge based on existing understandings. However, social constructivism adds a social dimension, considering elements like social interaction and language development as crucial means of learning (Retnowati, et al., 2016; Stahl, 2006).

Unlike cognitive constructivism, which posits that knowledge is individually constructed through personal processes, social constructivism asserts that knowledge is collaboratively constructed through social interactions, whether between peers or between teachers and students. Kalina and Powell (2009) distinguish cognitive constructivism from social constructivism based on Piaget’s and Vygotsky’s theoretical perspectives, respectively. They highlight key differences: "Piaget's theory places heavy emphasis on individuals' reasoning abilities and how they interpret knowledge. Vygotsky, on the other hand, believed that variables such as social interaction, culture, and language (communication) influence how individuals acquire knowledge" (p. 246).

The cognitive viewpoint asserts that learning involves internalization. Individuals construct internal mental representations by assimilating or accommodating new information to what they already know. Language and social interaction serve as channels to express the outcomes of these mental processes. Conversely, the social perspective contends that learning encompasses both externalization and internalization (Cress & Kimmerle, 2008). Language and social interaction act as primary tools for externalizing mental representations, as Aiello and Thurlow (2006) notes, "language doesn't mirror reality; quite the contrary, language constructs reality" (p. 93). Vygotsky suggests that communication is a prerequisite for social interaction and externalization, and that new mental representations are internalized during social engagement. Shared representations and co-constructed knowledge become essential components of collaborative learning.

2.3 Internalization and Externalization

Within the broad umbrella of social constructivism, collaborative knowledge construction can be understood as a dual process involving two cycles: personal comprehension and co-construction, which correspond to internalization and externalization, respectively. From a socio-cultural standpoint, individuals grasp understanding through the interpretation of their beliefs and perspectives. They then articulate their initial interpretations in words to participate in social processes, fostering the collaborative creation of new meanings (Stahl, 2006). The accumulation of initial discourse is refined and reinterpreted over time through negotiation, ultimately culminating in shared understanding (Stahl, 2006). As Stahl explains, "Beliefs transform into knowledge through social interaction, communication, discussion, clarification, and negotiation. Knowledge is a product mediated by social interaction." This model embodies the socio-cultural notion of externalization through the communication process to establish shared understanding and internalization through personal interpretation to give meaning.

The concepts of internalization and externalization for knowledge construction can also be expounded in other models. Cress and Kimmerle (2008) assert that social and cognitive systems, in a state of structural coupling, mutually evolve through language and communication. They contend that learning unfolds through both externalization and internalization, as individuals "cannot externalize their own knowledge without inducing changes in their individual knowledge" (p. 109). In this context, externalization triggers internalization and blurs the boundaries. They define this mutual development as the co-evolution of individual learning and the collective co-construction of group knowledge. Chen and Techawitthayachinda (2021) propose that social interaction enabling the externalization of uncertainty creates a necessary space for shared cognitive resources. The ability and opportunity to express one's uncertainty are initial steps to reduce uncertainty, progressively refining group knowledge through conversation, and shaping individual understanding.

2.4 Cognitive Load Theory and Multimedia Learning

Given that collaborative modes alter communication methods and information ecosystems (Cress, 2020; Smith et al., 2003), this section investigates how communication design and environments impact information processing, drawing on insights from cognitive load theory and multimedia learning. Cognitive load theory and multimedia learning theory are two main pillars that have parallelly developed from the same foundation of information processing theory to understand how communication environments affect individuals in managing the mental effort required for processing information.

Cognitive load theory (CLT) is based on the basic premise that human cognitive processing is constrained due to limited capacity of working memory. As the result, CLT have been developed to explain “how the information processing load induced by learning tasks can affect students' ability to process new information and to construct knowledge in long-term memory" (Sweller, 2019, p. 261). CLT is widely adopted in instructional design and education technology to optimize cognitive processing spent in material and learning activities.

Cognitive theory of multimedia learning (CTML) is grounded from three main cognitive principles: (a) dual channel (i.e., human possess two separate channels for visual and audio processing), (b) limited capacity (i.e., each channel has limited processing capacity), and (c) active learning (i.e., meaningful learning requires active learning, including select relevant information, coherently organizing selected information, and integrating incoming information with existing knowledge) (Mayer & Moreno, 2010). From CTML, students perceive either visual or auditory information through separate sensory memory (i.e., ears or eyes). These two kinds of information are transferred to working memory, mentally organized and integrated the information together with coordination from existing knowledge (Mayer & Moreno, 2010; Moreno, 2006).

Cognitive Effects from CLT and CTML

Since CTML is based on assumption that to engage in meaningful learning "people must actively process the incoming material in information processing channels that are highly limited" (Mayer & Moreno, 2010, p. 132), CTML shares several theoretical assumptions of cognitive processing with CLT (e.g., categories of cognitive load: Intrinsic, Extraneous and Germane cognitive load) and therefore shares some cognitive effects on learning. Thus, the relevant cognitive effects deriving from CLT and CTML may predict the quality of collaborative modes.

Modality Effect. Because working memory capacity is fixed and limited, the main mechanism under modality effect is to design material, instruction or environment in a way that shift information load from visual to auditory channels. This effect is based on assumption that working memory can be subdivided into two channels for processing visual and auditory information (Erladson et al, 2010; Nelson & Erladson, 2008; Sweller, 2022). Thus, when effectively distribute cognitive load from visual and auditory channels, working memory can process information more effectively (Sweller, 2022). One of the classic implications of this effect is found in multimedia design principle, suggesting designers to "present the words as concurrent narration rather than as concurrent on-screen text, thus off-loading the processing of the words from the visual channel to the verbal channel" (Mayer & Moreno, 2010, p.147).

Split-Attention Effect. Split-attention effect grew out of modality effect and complement each other. Generally, students need to mentally integrate many sources of information to arrive solution. However, if related sources of visual information are presented far apart (i.e., contiguity effect), it requires higher cognitive load to integrate all components by oneself than process already integrated information (Sweller, 2020). Another way to reduce split-attention effect can tie back to the modality effect in multimedia design principle that turning some visual information to be auditory information reduce split-attention effect and allows more space in visual channel to process visual information.

Collective Memory Effect. Based on the initiate analogy that learners are processors of information (Mayer, 1996), Kirschner et al. (2018) see collective memory as the product of larger, more effective processors. In collaboration, the processing power is divided among group members, which results in cognitive load reduction. These divided processors (i.e., learners) complement each other process information by filling the knowledge gaps (Sweller, 2020). However, there is cost of cognitive load from communication and coordination (i.e., transaction costs) to integrate the information elements. Therefore, one needs to consider the trad-off between the benefits of shared processing and transaction costs to decide the efficiency between group and individual learning (Janssen et al., 2010; Kirschner et al., 2018; Sweller, 2020).

Literature Review

This section delves into the challenges of video-based learning and examines the solutions and strategies discussed in previous literature. Subsequently, it goes on to explore Collaborative Video Viewing (CVV) in greater depth as an instructional strategy for promoting active learning. Later in the literature review, attention shifts to the affordances of temporality (synchronous and asynchronous) and modality (speaking and typing), as well as communication modes implemented in online CVV. Additionally, the review focuses on the learning opportunities presented by video annotation (an asynchronous text-based communication tool) and videoconferencing (a synchronous voice-based tool) within the context of CVV.

2.5 Video-based Learning

Lecturing videos serve as the primary mode of content delivery in many distance learning contexts. Online learning platforms like Khan Academy or MOOCs (massive open online courses) rely heavily on online videos, making experts and their expertise available on-demand, with scalable and cost-effective potential (Hansch et al., 2015;

Lemay & Doleck, 2020; Noetel et al., 2021; Stöhr et al., 2019). Video-based learning is also integral to flipped classrooms, where students independently learn from instructional videos and then engage in face-to-face discussions with peers (Lee et al., 2021). The benefits of video-based learning arise from the autonomy it offers, enabling students to pace their learning and tailor their environment to their cognitive needs (Costley et al, 2021; Schroeder et al., 2019; Schwan & Riempp, 2004). The visually rich learning environment is potent for visual strategies (seeing), vicarious experiences (doing), and motivation (engaging) (de Koning et al., 2018; Koumi, 2006). However, the passive nature of video viewing and the lack of interaction with fellow learners have faced criticism (Hew & Cheung, 2014; Hansch et al., 2015).

Strategies have been deployed to transcend knowledge transfer and facilitate knowledge construction, shifting student engagement from passive to active learning. One avenue to optimize learning from video lessons lies within the instructional videos themselves. For decades, principles of multimedia learning and cognitive load theory have guided effective multimedia design in instructional videos, primarily aimed at managing intrinsic load, enhancing germane load, and reducing extraneous load (Fiorella et al., 2020; Mayer & Moreno, 2010; Sweller, 2022). Instructional design principles have also been applied to enhance teaching methods, instructional sequences, presentations, and interface interactions in video lessons (Fiorella et al., 2020; Fiorella & Mayer, 2018; Ou et al., 2019).

Numerous variations of video formats have been introduced to mitigate passive learning in traditional talking-head, lecture-style videos, providing students the opportunity for more constructive engagement. For instance, within-video actions such as handwriting or drawing have been incorporated to enhance engagement and retention (Chen & Thomas, 2020; Fiorella et al., 2020; Kokoç et al., 2020). Dialogue tutoring videos have emerged to foster constructive learning engagement through vicarious tutoring (Chi et al., 2017; Muldner et al., 2014).

Another avenue for enhancing video-based learning extends beyond the videos themselves. Hansch et al. (2015) caution that crafting high-quality instructional videos necessitates deliberate instructional design, professional production, and time investment, resulting in significant expenses. Thus, educators need to critically evaluate the need for video production and explore alternative approaches that can achieve the same level of learning outcomes with reduced economic resources (Hansch et al., 2015).

Another strategy to curb passive learning while watching videos is to design active learning activities outside of the videos. Unlike creating active videos, integrating active learning activities can be a cost-effective strategy, requiring fewer video production skills and expenditures, as repurposing existing videos is more cost-efficient than creating new ones (Brame, 2017; Hansch et al., 2015). An investigation into what types of learning activities to incorporate and how to integrate them into video-based learning across diverse domains of knowledge is imperative for advancing the quality of video-based learning (Caspi et al., 2005; Fiorella et al., 2020; Kaiser & Mayer, 2019; Muldner et al., 2011).

2.6 Collaborative Video Viewing (CVV)

Collaborative Video Viewing (CVV) is an educational activity wherein pairs or groups of students work together to accomplish learning tasks while simultaneously watching a video. These tasks may involve discussions, concept explanations, or problem-solving. CVV injects a social dimension into video-based learning and leverages collaboration to foster social engagement, enabling students to delve into more profound instructional content and active learning.

The concept of CVV originated from blended learning, which integrates online educational materials into traditional face-to-face classrooms to stimulate discussions, reflections, and an enhanced learning experience (Gibbons et al., 1977). Across diverse learning contexts, students have demonstrated improved learning outcomes and expressed greater satisfaction with collocated synchronous watching compared to independent viewing (Pi et al., 2020; Li et al., 2014).

With the advancements in internet technology, the concept of distributed collaborative video observing, or online CVV, has been developed to replicate the advantages of face-to-face CVV within the realm of online learning. The idea of "watching together" in the context of online learning extends beyond the confines of co-location (location dimension) and synchronous viewing (time dimension). It also encompasses asynchronous viewing, provided that students are watching the same video with shared cognitive focus on learning objectives or tasks (objective dimension). In the realm of online learning, CVV enriches both social and cognitive engagement. The introduction of a social aspect into video viewing converts individual activity into a collective endeavor, consequently amplifying the potency of interventions at the individual level (Liu et al, 2023).

On one hand, the benefits of collaborative viewing stem from increased social engagement. Prior studies have revealed that students tend to lose focus and engage in mind-wandering for almost half the duration of lengthy lecture videos (Kane et al., 2017;

Szpunar et al., 2013; Zhang et al., 2020). Integrating social interactions into the video viewing experience serves as a strategy to minimize mind-wandering and keep students attentive. Morse (2021) found that engaging in conversation with peers reduces mind-wandering to about a quarter of the time. Additionally, analysis of human-computer interaction has shown that students often skip the beginning and end of videos, concentrating primarily on lecture content while ignoring instructor comments (Chen et al., 2016). Nevertheless, engaging collaboratively with peers can instigate peer influence that motivates students to uphold a sense of responsibility, remain focused, and sustain motivation during co-watching videos (Liu et al., 2023; Schneider & Pea, 2013). When their peers are concurrently watching the same video, students might exhibit reluctance to skip any content and even need approval from their partners. Even in asynchronous watching scenarios where students have more autonomy over video control, collaborative annotation tools have proven highly motivating due to the influence of social connections, as evidenced by dynamic text dialogues (King & Sen, 2013).

On the other hand, from a cognitive perspective, the advantages of collaborative viewing stem from active participation in joint learning efforts. Grounded in the ICAP framework (Chi & Wylie, 2014), even superficial engagement in conversations—such as expressing agreement ("right," "yeah") or reiterating existing material—contributes to information processing reinforcement and the activation of relevant schemas tied to the ongoing learning task. The opportunity to express personal opinions can trigger a constructive process wherein students revise or expand their understanding, leading to knowledge construction. In the ideal scenario, CVV enables students to leverage social interactions to build upon each other's ideas and collaboratively construct knowledge, a feat often unattainable through individual observation.

2.7 CVV Collaborative Modes

In the realm of collaborative video viewing (CVV), the current implementations have been limited to tools that utilize synchronous voice and asynchronous text communication. Theoretically, comparing these current modes may introduce confounding factors between temporality and modality. Nevertheless, a direct examination of these two prevalent collaborative modes proves to be a feasible approach for practical implementation in active video-based learning. In this section, this review delves into the potential rationale for incorporating asynchronous text-based and synchronous voice-based collaborative modes in the context of CVV activities. Furthermore, the review explores the benefits of such collaborative modes and tools, drawing upon existing literature.

The motivation for including asynchronous text-based and synchronous voice-based collaborative modes, while not incorporating the other two modes in previous literature, may be elucidated by cognitive load theory within multimedia design. For instance, the absence of tools employing asynchronous voice-based modes in CVV could be attributed to the potential cognitive overwhelm stemming from the integration of multiple multimedia elements. Asynchronous voice-based tools such as VoiceThread usually require students to record videos or use their voice for responses. This technology has a positive impact on enhancing social presence and makes it easier for students to generate a substantial volume of learning outputs compared to asynchronous text-based discussions (Lowenthal et al., 2020). However, in the context of CVV, where the central point of discussion is instructional videos, introducing additional competitive media elements, like playing video or audio clips shared by peers, could potentially overwhelm students with cognitive load. This extraneous cognitive load stems from the requirement for students to adeptly manage multiple multimedia components concurrently displayed on the screen (Dirkx et al., 2021; Mayer, 2021; Sweller, 2020).

However, in asynchronous text-based tools, the control of media is less burdensome because these collaborative tools separate the communication modes and control between video and collaboration. While asynchronous voice communication demands interface interaction (i.e., click to play or pause) to enable the partner's voice, which is the same action as video control, the persistent text affordance in textual modality does not require interface interaction to access information and thus allows students to shift their attention to interface control for video control.

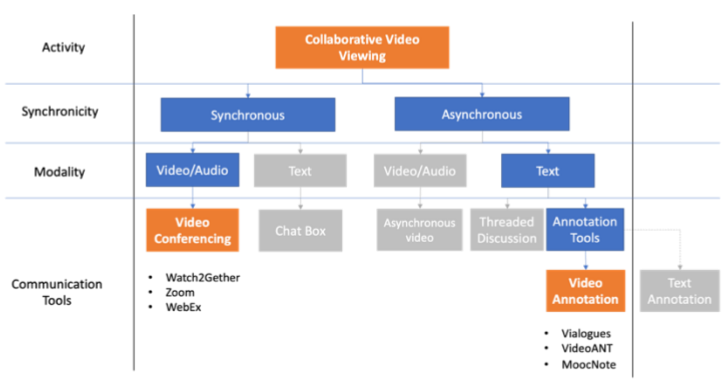
Similarly, in their comparative study of asynchronous and synchronous text-based tools for collaborative video viewing, Cadiz et al. (2000) observed that participants using synchronous text-based tools reported experiencing adverse perceptions of cognitive load and split attention. This was attributed to the simultaneous dual-screen communication, involving both a text screen and a video screen, while utilizing instant messengers to communicate while watching videos. Participants reported the need to split their attention between the video and chat window simultaneously. This visual load, stemming from processing information from two visual sources concurrently, overwhelmed them as they tried to keep up with both video content and chat messages.

However, the adverse effect of dual-screen communication was less pronounced in the case of synchronous voice-based communication, where participants could utilize two separate visual and audio channels for communication (e.g., Cadiz et al., 2000; Erlandson et al, 2010). The separation of visual-audio channels aligns with the modality principle, which posits that individuals acquire knowledge more effectively when information is delivered through audio narration instead of on-screen text (Mayer, 2005). By shifting text content from the visual channel to the verbal processing channel, it can help mitigate the adverse split-attention effect, where learners constantly switch their focus between different related sources of information (Moreno, 2006).

Due to the control and visual challenges posed by asynchronous voice-based and synchronous text-based modes respectively, the CVV context primarily accommodates the application of asynchronous text-based (AT) and synchronous voice-based (SV) tools (see Figure 1). Among the communication tools available in both modes, video conferences and video annotation tools have been thoroughly examined and developed, gaining widespread use in various CVV settings across diverse disciplines. However, there exists a notable gap in CVV research, as no study has directly compared the learning processes and outcomes resulting from the utilization of tools that give rise to this distinctive communication and learning environments.

Figure 1

Collaborative Modes and Tools



2.7.1 Asynchronous Text-based Collaborative Mode (AT)

Online communication primarily relies on asynchronous text-based computer-mediated communication (Hewitt, 2005). This form of communication transcends spatial and temporal constraints, enabling students to engage at their preferred times and locations. In addition to its practical advantages, asynchronous text-based communication (AT) offers several cognitive benefits stemming from its time flexibility, delayed interactions, and written format.

Asynchronous text-based communication offers students the freedom to think and generate intricate, thoughtful, and contemplative responses without disruptions from others (Griffiths & Graham, 2010; Hrastinski, 2008). A survey on students' perceptions indicated a preference for asynchronous text-based (AT) communication over face-to-face communication when engaging in brainstorming tasks. This preference stems from the ample time it provides for reflection and idea development. (An & Frick, 2006). Despite their positive views on the opportunity for reflection, many students tend to limit their participation to meeting the minimum posting requirement and do not engage in profound reflection on the content or follow up on their peers' contributions (Andresen, 2009; Choi & Hur, 2023).

Asynchronous text-based communication not only enhances students' reflective thinking but also aids in organizing their thoughts for more effective communication. In AT, students can strategically allocate cognitive resources to plan, compose, edit, and review their messages before delivery, resulting in more deliberate, concise, and accurate outputs (Hew & Cheung, 2013; Sherblom, 2010; Stockwell, 2010). Beyond these benefits, it may be particularly advantageous for individuals with certain personality traits, as it reduces the feeling of being "put on the spot" or facing social presence demands (An & Frick, 2006; Johnson, 2006; Kehrwald, 2008).

Asynchronous text-based collaborative tools not only provide individuals with a persistent record of group interactions and collaboration progress but also serve as a shared information repository, fostering collaborative group work (Jeong & Hmelo-Silver, 2016; Suthers, 2006). AT discussions establish repositories of collective knowledge, facilitating future references, revisions, and study reviews (Andresen, 2009). The written materials within these platforms become fundamental components of the group's knowledge, functioning as historical records of negotiations and implicitly documenting changes and rationales within the knowledge-building community (Jeong & Hmelo-Silver, 2016; Scardamalia & Bereiter, 2003; Stahl, 2003; Suthers, 2006).

2.7.2 Synchronous Voice-based Collaborative Mode (SV)

Synchronous voice-based communication (SV) is relatively less common in online environments compared to asynchronous text-based modes due to its inherent time constraints (Watts, 2016). Nevertheless, it can offer distinct advantages for specific learning tasks and learner groups, particularly those characterized by age and competency levels, where immediate group interaction and spoken communication prove beneficial (Grabowski, 2010).

Gunawardena (1995) emphasized that communication serves the dual purpose of conveying both message content and social information. In group interactions, individuals extract social cues and nonverbal signals to reduce communication ambiguity and establish positive impressions (Tanis & Postmes, 2003). Drawing from rich media theory, tools equipped with audiovisual features, such as videoconferencing, provide personal information including tone and social cues, enabling quick clarification of ambiguity and facilitating effective conversation (Hew & Cheung, 2013; Peterson & Roseth, 2015). Social information enriches contextual understanding, reducing interpersonal uncertainty and aiding in tasks that involve ambiguity, such as decision making, relational development, and problem-solving (Griffiths & Garnham, 2010; Sherblom, 2010).

Gunawardena (1995) underscored the dual role of communication, emphasizing its function in conveying both message content and social information. On the one hand, social information enhances contextual comprehension, reduces uncertainty in interpersonal interactions, and assists in ambiguous tasks requiring convergence, such as decision-making and problem-solving (Griffiths & Garnham, 2010; Sherblom, 2010). On the other hand, social information in communication fosters social engagement, which profoundly impacts both individual generative processes and the cognitive development of the group. Pi et al. (2022) discovered that subtle social cues, such as nodding in approval, enhance the generative process by providing social information in the form of positive feedback.

Synchronous voice-based communication facilitates a turn-taking dialogue style that resembles face-to-face interaction. Chi and Wylie (2014) asserted that interactive behaviors require "a sufficient degree of turn-taking" (p. 223) to collaboratively develop hypotheses, negotiate perspectives, and build shared understandings. However, a mere quantity of turn-taking isn't sufficient for effective co-construction; the quality of co-construction relies on the number of co-constructive turns in dialogue (i.e., Interactive in

ICAP), reflecting contributions that build upon previous ones, rather than mere acknowledgment or repetition (i.e., Active in ICAP; Chi & Wylie, 2014; Chi et al., 2017).

Immediate responses in synchronous voice-based communication serve to reduce cognitive conflict and ambiguity upfront, facilitating continuous knowledge construction (Hrastinski, 2008; Stahl, 2003; Sweller, 2019). Drawing from cognitive load theory, immediate interactions aid students in offloading their cognitive load since processing power is distributed among group members, and individual memory transforms into collective memory accessible to all (Kirschner et al., 2018). This immediacy of verbal modality in synchronous voice-based communication also reduces communication filtering, motivating students to express their ideas and engage in the generative process. Instructors in Park and Bonk's study (2007) observed a distinction between synchronous voice-based and asynchronous text-based communication, noting that "students are very cautious and conservative in the amount of what they say or what they try to address in an asynchronous discussion forum. However, synchronously, especially with voice, they go faster, and they try things out a little more" (p. 251).

While co-construction does not necessarily demand immediacy, immediate responses enable students to dynamically evolve their ideas through iterative processes involving both low-level (e.g., more detailed and concrete) and high-level (e.g., more abstract and conceptual) concepts. Dillenbourg (1999) argued that when students rephrase what others have said, either providing more detail or presenting concepts, they address misunderstandings and reformulate prior statements in a more elaborative and coherent manner, ultimately leading to the reconstruction of mental models.

2.8 CVV Collaborative Tools

2.8.1 Collaborative Video Annotation in CVV

Online video annotation tools are web-based applications designed for tasks like clipping, segmenting, or marking specific portions of videos. These marked segments can include timestamps and visual elements accompanied by text and visual comments (Sablić et al., 2021; Lam & Habil, 2021). Video annotation tools find extensive use in collaborative video annotation, where multiple individuals collectively watch and annotate videos for shared tasks (Chan & Pow, 2020).

These tools are widely employed in fields such as education and medicine to observe and analyze observable behaviors in instructional videos, aiding in practice reflection, self-efficacy, and skill development. Examples of contexts where video annotation tools are applied include improving teaching practices, enhancing clinical skills, and developing workplace and interpersonal abilities (Girasoli & Hannafin, 2008; Lam & Habil, 2021; Rich & Hannafin, 2009). While some studies explore how collaborative video annotation tools enhance conceptual understanding of video content, only a limited number investigate their potential implications for co-constructing knowledge and facilitating knowledge transfer (Evi-Colombo et al., 2020, Novak et al., 2012).

Before the introduction of annotation tools in Collaborative Video Viewing (CVV) contexts, traditional asynchronous text tools used for CVV typically involved threaded discussions. In these discussions, students would independently watch videos and respond to question prompts related to the video content (Sherer & Shea, 2011). However, these text-heavy asynchronous forums could overwhelm students, disrupt their thought processes, and hinder collaborative knowledge construction (Andresen, 2009; Sun & Gao, 2017). Hewitt (2005) found that students often adopted a "single-pass" strategy, reading only new posts and neglecting previous ones, causing threaded discussions to prematurely lose momentum.

To address these issues, annotation tools were introduced to reduce text overload and help students stay focused on key ideas (Sun & Gao, 2017). Over time, the use of annotation tools expanded to encompass not just textual information but also video annotations as technology advanced. Various forms of collaborative video annotation tools have been integrated into CVV to promote active learning. Students participating in annotated discussions tend to engage with specific content sections, leading to more focused and meaningful contributions compared to traditional threaded discussions (Chan & Pow, 2020; Morales et al., 2022). The reduced volume of annotations also lowers cognitive load and facilitates follow-up group conversations (Agarwala et al., 2012; Kawase et al., 2011). Annotated threads often adopt a turn-taking style, emulating dialogue instead of essay-style responses in threaded discussions, fostering active video-driven conversations rather than passive video viewing (Hsiao et al., 2014). This immediacy adds a social element of "watching together" to asynchronous text-based interactions.

However, despite technological advancements in video annotation tools, these asynchronous text-based tools may still hinder students from promptly building upon each other's ideas, resolving ambiguity, and addressing misunderstandings, potentially impeding the continuity of the knowledge co-construction process (Hewitt, 2005; Hrastinski, 2008; Stahl, 2003). Some students primarily use video annotation tools for personal bookmarking and memorization or to share highlighted information, rather than fully leveraging the social component of the tool for interactive knowledge co-construction (Chan & Pow, 2020; Kawase et al., 2011; Novak et al., 2012; Shelton et al., 2016).

2.8.2 Video Conferencing in CVV

In contrast to the growing popularity of online synchronous collaboration in the past decade, a few studies up to date implement synchronous voice-based collaborative modes (SV) in online educational CVV. One of the barriers is the synchronous requirement that poses time constraints. Another technical barrier is that synchronous voice-based collaboration in online CVV can reduce a student's learning autonomy due to shared video control and can lead to a higher cognitive load from video control coordination (i.e., transaction costs) (Cho & Lee, 2013; Kirschner et al., 2018; Sun et al., 2017; Sweller, 2020).

Previous videoconference tools did not allow students to watch online videos together while everyone had real-time control over the same video (Cadiz et al., 2000). In an online CVV, one participant usually hosts and controls the video (e.g., play and pause) and shares the screen with others (Sun et al., 2017). Viewing coordination between a host and a participant can occur when a participant wants to discuss or comment at any point in the video and wishes to go back or forward the video to focus on a particular section. These situations require a participant to communicate their need and coordinate with a video host to pause or jump to requested points. However, this coordination may have drawbacks as the coordination cost becomes more significant in an online collaboration where several social cues and emotional connections are limited. Some participants in online CVV may try to skip their pause-to-comment because they feel that it interrupts other participants or simply avoid it as a strategy to cope with coordination costs (Geerts, 2006).

However, many web platforms have recently enabled synchronized social viewing, incorporating features like shared navigation control commands and audiovisual channels for group interaction and collaboration, as seen in platforms such as Newrow, Synchtube, and Wersync (Boronat et al., 2021). These synchronized social viewing platforms may compromise students' learning autonomy but provide greater opportunities for social interaction in online CVV.

A significant body of literature on human-computer interaction in co-watching activities focuses on the user experience with different technology configurations, including collaborative modalities. For instance, Geerts et al. (2011) found that co-watching activities increase the sense of togetherness and establish an immediate common ground for participants to discuss while watching synchronized videos. In their study on synchronous CVV, participants generally preferred the voice condition over the text condition. They noted a preference for the natural, direct contact, and intuitive nature of audio-based communication. Conversely, they reported a higher cognitive load in text-based communication, particularly when they needed to divide their attention between the video and the chatbox, especially when pausing control was not available.

Cadiz et al. (2000) examined how various collaborative modes affected the online CVV experience in educational contexts, considering both the learning process and outcomes. Similar to Geerts et al. (2011), they found an adverse effect on cognitive load due to split attention when multitasking and engaging in two-screen viewing with a chatbox. Nevertheless, when comparing face-to-face CVV to online CVV using a videoconferencing application, there was no difference in learning outcomes, as comprehension quiz scores remained the same in both conditions.

Hypotheses

With the growing prominence of online education, video-based learning has gained increasing traction as a means to engage and educate learners. Collaborative video viewing holds the promise of transforming passive video watching into an active learning experience. This study seeks to explore and compare the learning engagement and interactions of participants in synchronous voice-based and asynchronous text-based collaborative viewing modes to shed light on the extent and nature of the impact of collaborative modes on learning outcomes in an online setting.

Building upon the findings of previous literature reviews, multiple studies have shed light on the facilitation of synchronous voice-based collaborative modes in the domain of transferable problem-solving. Without group interaction, oral explanations have demonstrated a greater capacity to effectively stimulate students' elaborative cognitive processes when compared to written explanations, resulting in a more significant advantage in the acquisition of transferable knowledge (Lachner et al., 2018). Furthermore, immediacy within communication has been shown to reduce cognitive conflicts and ambiguities, which play a crucial role in promoting negotiation and the continuous co-construction of knowledge among group members (Hrastinski, 2008; Stahl, 2003; Sweller, 2019).

From a social perspective, social information, such as social cues from verbal modality, enhances contextual comprehension and reduces uncertainty in interpersonal interactions. Other affordances that are strengthened as a result of the interaction between immediacy and verbal modality include social presence, leading to increased social engagement and motivation. Similarly, immediate verbal communication also minimizes communication filtering, encouraging students to freely express their ideas and actively engage in the generative process (Park & Bonk, 2007).

This study hypothesizes that synchronous voice-based collaboration motivates learners to be more elaborative, resulting in better transfer scores. However, the immediate interaction in the synchronous dimension, which creates a more interactive dynamic, tends to favor interactive processes over constructive processes. The dynamic discussion in synchronous voice-based collaboration also supports clarification and shared understanding in the negotiation process, thereby reducing the learning gap among members. The following hypotheses outline our expectations regarding participant performance and collaborative behaviors in these two distinct modes.

Hypothesis 1: Individuals participating in synchronous voice-based collaborative viewing (SV) are expected achieve higher scores compared to those engaged in asynchronous text-based collaborative viewing (AT). Furthermore, it is anticipated that synchronous voice-based dyads will exhibit performance consistency within the dyad.

Hypothesis 2: Dyads involved in synchronous voice-based collaborative viewing (SV) are anticipated to exhibit a greater number of interactive comments, while those engaged in asynchronous collaborative viewing (AT) are expected to demonstrate a higher frequency of constructive comments. Based on ICAP, within the same mode, dyads with higher scores will likely display a greater number of generative comments, as compared to dyads with lower scores.

Hypothesis 3: Dyads participating in both collaborative modes are expected to engage in the co-construction process, with dyads in synchronous voice-based collaboration exhibiting a higher degree of negotiative discourses.

CHAPTER 3

METHODOLOGY

Methods

The objective of this study is to compare the impact of two different modes of collaborative watching, namely synchronous voice-based mode (SV) using Watch2Gether and asynchronous mode (AT) using VideoAnt, on the learning process and outcome. By investigating the following three research questions, this study aims to provide valuable insights into the collaborative learning process, guiding instructional designers and educators in their decision-making regarding instructional tools, video-based learning, and the overall design and instructional support of online learning environments.

Table 1

Research Questions, Data Sources, and Analysis Methods

RQ	Research Questions	Data Source	Analysis
1	Are there differences in learning outcomes, between individuals and score-group dyads in synchronous and asynchronous collaborative watching?	Quiz scores	ANCOVA, K-Mean
2	What are the differences in learning engagement, between individuals and score-group dyads in synchronous and asynchronous collaborative watching?	Discussion	Content Analysis (ICAP), Independent t-test
3	What kind of discourse and discourse patterns do score-group dyads employ to co-construct knowledge in synchronous and asynchronous collaborative watching?	Discussion	Content Analysis (Interactive Discourse), Bigram, Sankey Diagram, Thematic Narrative Analysis

3.1 Participants

The author distributed online and printed research recruitment advertisements targeting undergraduate students at a university in the southwest region. The participants were required to be over the age of eighteen and represented various majors and academic years to ensure a diverse representation of the undergraduate population. There were no specific prerequisites or prequalifications for participation in the study, except for having access to a computer and the internet to partake in the experiments.

3.2 Research Design

Undergraduate students were recruited to participate in two online video sessions, utilizing a synchronous collaborative watching mode (Watch2Gether) and an asynchronous collaborative watching mode (AntVideo), to watch videos and engage in discussions based on prompts.

Employing a within-subject, counterbalanced experimental design, the first session involved forty participants divided into two groups. The first group of 10 dyads (n = 20) used Watch2Gether (synchronous voice-based mode) to watch a pricing-bias video, while the second group of 10 dyads (n = 20) used AntVideo (asynchronous text-based mode). In the second session, the groups and conditions were switched, with the first group using AntVideo to watch a perception-bias video and the second group using Watch2Gether for the same video. However, four participants dropped out after the first session. Therefore, in the second experiment, the first 9 dyads (n = 18) utilized AntVideo (asynchronous mode) to watch a pricing-bias video, and the remaining 9 dyads of the second group (n = 18) used Watch2Gether (synchronous mode) for the same video. As a result, this study involved 40 participants in the first experiment and 36 participants in the second experiment, resulting in a total of 76 data points (N = 76). These data points were divided into 38 from the asynchronous mode (n_{asyn} = 38) and 38 from the synchronous mode (n_{syn} = 38).

A within-subject design was employed in this study, where all participants experienced the same conditions using collaborative watching tools, allowing for the measurement of changes resulting from different treatments. This design, with the same number of participants, holds greater statistical power compared to a between-subjects design as it eliminates individual variation. Nonetheless, a notable limitation of the within-subject design is the potential for carryover effects, where participation in one condition may influence performance or behavior in all subsequent conditions. To address this concern, the study employed counterbalancing to ensure a balanced presentation of the sequences of collaborative watching modes.

3.3 Communication Modes and Tools

The affordances inherent in collaborative tools are a result of their collaborative modality and specific design features. In this study, Watch2Geter is selected as the synchronous voice-based collaborative tool, while VideoANT is chosen as the asynchronous text-based collaborative tool. These selections were made to investigate how distinct collaborative modalities and tool functionalities impact the collaborative process and learning outcomes.

Watch2Geter: Synchronous Voice-based Collaborative Tool (SV)

Watch2Gether is an online web application that operates synchronously, allowing instructors to establish a virtual room that promotes collaborative video co-viewing and cultivates a social sense of shared watching. Once provided with a link, students can access the video and collectively control its playback. Simultaneously, they have the option to engage in video conferencing using microphones and cameras, facilitating discussions related to the video content and enhancing the social connectedness during the shared viewing experience. Watch2Gether offers two distinct advantages in comparison to other conferencing programs. Firstly, it eliminates the need for students to install additional plugins or applications, thereby accommodating individuals with limited computer literacy or those utilizing public computers. Additionally, each student maintains full control over the video independently (e.g., play, pause, and playback), reducing the transaction costs associated with video control and enhancing learner agency (Kirschner et al., 2018; Schroeder et al., 2020).

VideoAnt: Asynchronous Text-based Collaborative Tool (AT)

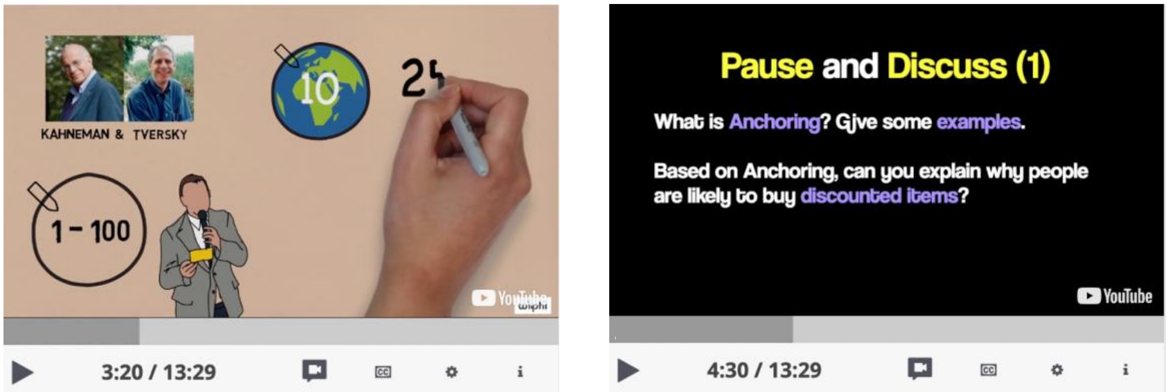
VideoAnt is an asynchronous video social annotation tool that integrates video annotation and discussion, with the aim of creating a community of learners who engage deeply with instructional content through video-based discussions. Unlike some other annotation tools that necessitate an instructor to initiate the post, this tool empowers participants with autonomy and self-direction (Van der Westhuizen, 2015). It enables them to annotate specific time points in a video and initiate discussion threads linked to each annotated timestamp, thereby replicating the dialogue-style interactions typically found in face-to-face settings. Hsiao et al. (2014) assert that "discussing around videos includes more complex interactions rather than having dialogues alone/among groups or merely performing video annotation" (p. 363). Participation in AntVideo involves the annotation of videos, leading to the formation of discussion threads that encompass titles, content boxes, and a submit button. Furthermore, students have the opportunity to engage in a reciprocal manner, using the response feature to respond to and expand upon the contributions made by their peers.

3.4 Instructional Materials

The experiments involved the presentation of two videos focusing on pricing and perception biases. Rigorous control was applied to video production to ensure that both videos maintained a consistent format and content structure. Specifically, the videos followed a narrative-over-visualization/hand drawing format, with a typical duration of approximately 12-13 minutes (See Figure 2). Each video was structured around three primary concepts pertaining to cognitive biases, all delivered by the same instructor. In the first experiment, the pricing-bias video expounded upon the Anchoring Effect, Pricing Biases, and Loss Aversion. Correspondingly, the second experiment featured a perception-bias video that explored Alief, Mental Accounting, and Peak-End Biases. To introduce each concept, the narrator commenced with a real-life story, illustrating the manifestation of the respective bias within practical scenarios. Subsequently, the presenter reinforced each concept by providing approximately four to five illustrative examples, facilitating a comprehensive understanding of the key ideas being conveyed.

Figure 2

Instructional Video Screen and Prompts



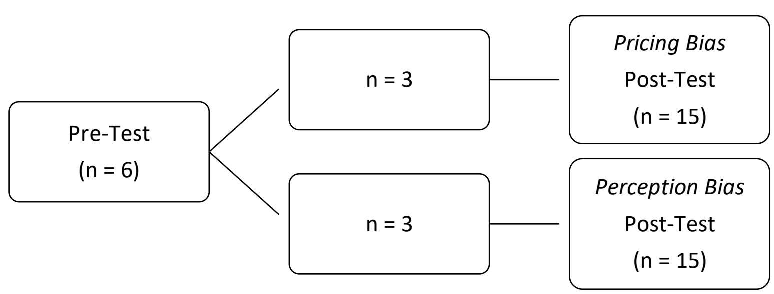
Following the explanation of each concept, prompts were presented to the dyads for approximately ten seconds. The screen displayed an instruction to pause the video in order to engage in discussion with their partner or write a discussion post/response, indicated as "Pause and Discuss" (See Figure 2). These prompts were meticulously designed to foster active participation and collaboration during the co-construction process. The prompts for each concept followed a consistent structure, aiming to elicit both conceptual understanding and practical application. For instance, the question "What is Alief? Give some examples" aimed to prompt students to provide a definition or demonstrate a fundamental understanding of the target concept. On the other hand, the question "Can you think of a situation when Alief is considered bad or good?" encouraged students to engage in critical thinking beyond the scope of the video content. The author deliberately crafted these prompts to align with assessments that measure knowledge retention and application. The underlying hypothesis being tested posits that groups with higher levels of engagement in discussion would exhibit superior learning outcomes.

3.5 Assessments

Prior to participating in the first experiment, each participant completed a pretest, followed by a posttest after watching each video. The pretest consisted of 6 questions randomly selected from 30 questions in the post-tests, with 3 randomly selected from a pool of 15 questions related to the pricing biases video, and the remaining 3 from the 15 questions pertaining to the perception biases video. Each question provided four options that corresponded to the concepts covered in the respective video (See Figure 3).

Figure 3

Pre-Test Structure

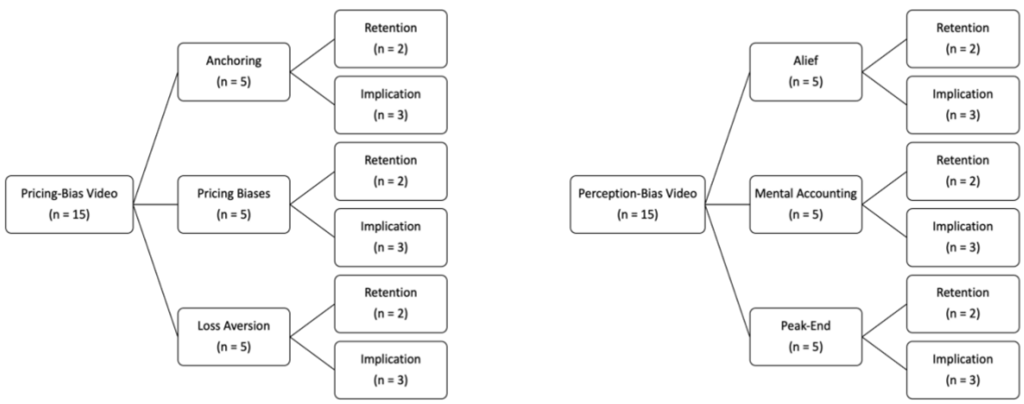


At the end of each co-watching session, participants underwent a multiple-choice quiz comprising fifteen questions. This quiz consisted of six retention questions and nine application questions. Within each video, the two retention questions and three application questions were derived from a single main concept (See Figure 4). The retention questions aimed to assess students' ability to recall information from the video and understand the fundamental ideas of the learned concepts. Examples of retention questions include “What is NOT an example of Alief based on philosopher Tamar Gendler?” and “Surprisingly, people feel less pain when having a longer painful experience because __”. Conversely, the application questions gauged participants’ capacity to apply the learned concepts in novel contexts or to develop a deeper understanding beyond the video content. Examples of application questions include "Which bias is most closely associated with Alexa's obsessive-compulsive disorder (OCD) behavior, where she constantly checks her bank account after making each purchase?" and "Based on the chart, rank the traumatic experiences of four individuals involved in a war in decreasing order."

3.6 Procedures

Figure 4

Post-Test Structure

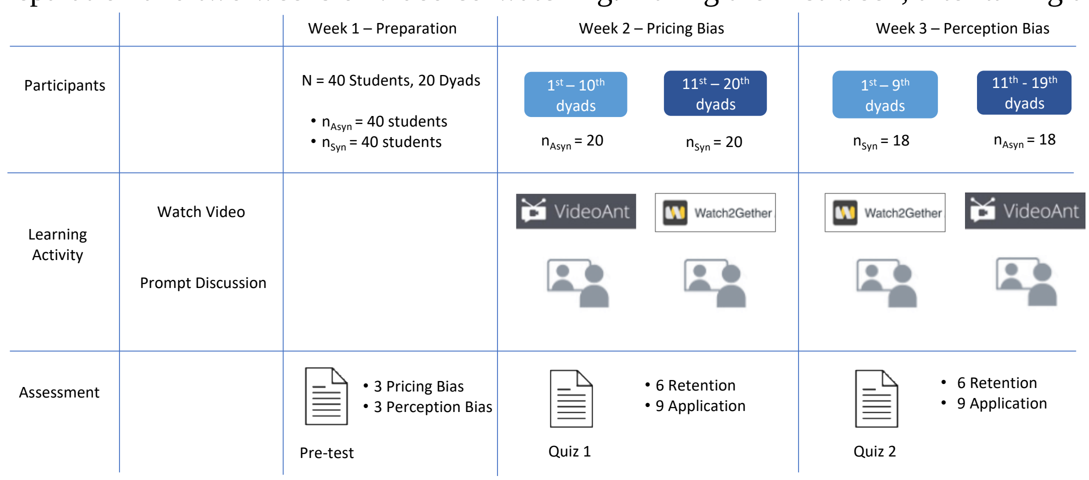


Undergraduate students were recruited to participate in two online sessions utilizing both synchronous voice-based (Watch2Gether) and asynchronous text-based (VideoAnt) collaborative watching tools. These sessions involved watching videos and engaging in mandatory prompt discussions inserted after each concept presented in the videos. A within-subject counterbalanced experimental design was employed, with 10 dyads assigned to watch a pricing-bias video using Watch2Gether and another 10 dyads assigned to watch a video using AntVideo in the first session. The groups and conditions were then switched in the second session. Following the first session, a few participants dropped out of the experiment. Consequently, in the second session, the first 9 dyads watched a perception-bias video using AntVideo, while the remaining 9 dyads watched the same video using Watch2Gether (See Figure 5).

The experiment lasted for approximately three weeks, encompassing a week of

Figure 5

Experiment Procedures



preparation and two weeks of video co-watching. During the first week, after taking a pre-test, the participants were randomly assigned to either group 1 or group 2. Group 1 was assigned to use VideoAnt (See Figure 6), whereas group 2 was assigned to use Watch2Gether (See Figure 7) for collaboratively watching a pricing-bias video.

In group 1, participants were randomly paired into 10 dyads. At the end of the first week, Group 1 received an experiment instruction via email, which included information about the learning activities, experiment timeline, a co-watching link, and setup tutorials for using VideoAnt. The dyads were instructed to watch the video, respond to video prompts, and discuss the video content at their own preferred times from Monday to Saturday. To attend the discussions, the dyads were required to access the assigned link and create video annotations. Furthermore, they were encouraged to revisit the video annotation thread to respond to and expand upon their partner's comments. The video annotations were securely stored on a server and later exported for analysis.

Group 2, participating in Watch2Gether, received a co-watching booking sheet to schedule their available time for watching the pricing-bias video during the first week. Based on the provided availability from Monday to Saturday, the author randomly paired the participants into 10 dyads. In the second week, once a dyad was successfully paired, the experiment instructions were sent to them via email, including details about the learning activities, experiment timeline, co-watching time, co-watching link, and Watch2Gether setup tutorials. The participants in Watch2Gether synchronously watched the video, responded to video prompts, and discussed the video content, while an author remotely recorded their shared screen and conversations. On Sunday of the second week, a quiz on pricing-bias concepts was administered to both experiment groups.

During the third week, the groups of participants and communication modes were switched. Group 1 was assigned to use Watch2Gether, while group 2 was assigned to use AntVideo for collaboratively watching a perception-bias video. The experiment procedures and timelines in the third week remained the same as in the second week. The dyads within each group were randomly re-matched. For example, in group 1, participant A, who was paired with participant B to watch the pricing-bias video on AntVideo in the second week, would be paired with participant C to watch the perception-bias video on Watch2Gether in the third week.

There were no restrictions on the amount of time spent watching videos in Watch2Gether or the number of annotations posted in AntVideo. Participants in Watch2Gether typically spent approximately 20 to 35 minutes watching videos and discussing prompts and content, while participants in AntVideo made around 3 to 25 video annotations.

Figure 6

VideoAnt: Asynchronous Text-based Collaborative Mode

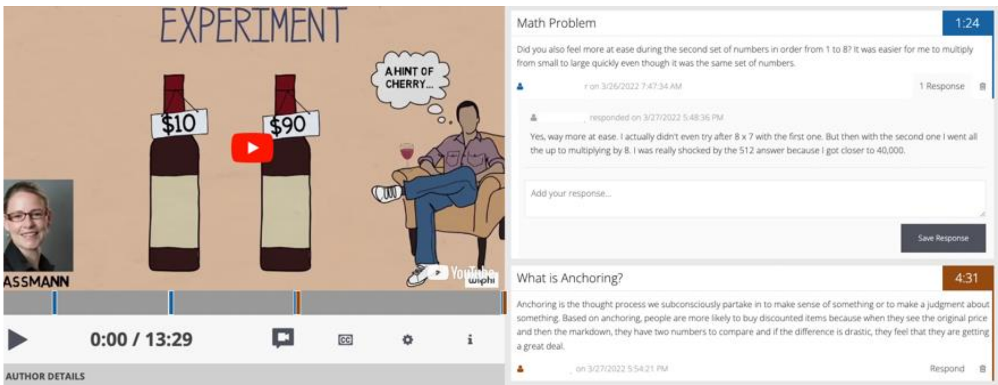
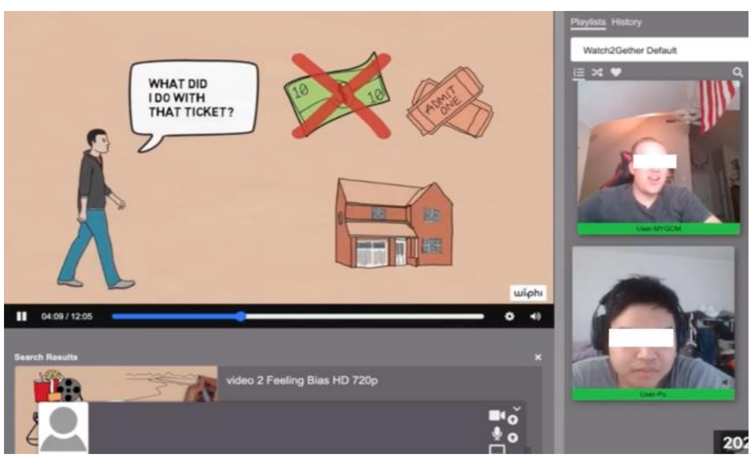


Figure 7

Watch2Gether: Synchronous Vext-based Collaborative Mode



Analyses

Research Question 1

3.7 Learning Outcomes

Outlier detection was conducted using SPSS Boxplots to identify extreme values in the sum, retention, and application scores for each video. SPSS applies the criterion of values falling outside the range of 3rd quartile + 1.5 times the interquartile range to identify outliers (Field, 2013). In the pricing biases video, the Boxplot analysis revealed that two participants displayed extreme-low outlier scores in all three categories (sum, retention, and application), while one participant exhibited an extreme-high outlier scores in the sum and application scores. In the perception biases video, one student demonstrated extreme-low outlier scores in the sum and retention categories. The primary aim of the outlier detection method employed in this study was to minimize the impact of discordant data resulting from inherent variations within the population or process, by considering the distance and density of data points (Salgado et al., 2016, p. 163).

A one-way between-group analysis of covariance (ANCOVA) was conducted to examine the differences in mean post-test scores within the sum, retention, and application scores among participants in asynchronous and synchronous collaborative modes for both videos. The independent variable in this ANCOVA analysis was the type of collaborative mode (synchronous and asynchronous), while the dependent variable consisted of the participants' scores in the post-tests for the sum, retention, and application categories. Covariates, including the participants' pretest scores and approximate GPA, were included to account for their prior knowledge and academic achievement prior to the treatment. To ensure uniformity and comparability, all scores were converted to percentages before conducting the analyses.

Following this, further one-way ANCOVA tests were performed to examine the mean differences in the sum, retention, and application scores between the asynchronous and synchronous collaborative modes within each video, specifically the pricing-bias and perception-bias videos. While controlling for the formats and difficulty levels in both videos, conducting separate ANCOVA tests allowed for the detection of any variations that might arise due to differences between the videos.

3.8 Dyad-Score Groups

Shifting our focus from individual to dyad levels of analysis, this study conducted K-means cluster analysis on the post-test scores of each dyad. This analysis aims to explore the relationship between collaborative modes, dyad performance, and dyad dynamics. The outcomes of the cluster analysis provided insights into the characteristics of dyad groups based on their performance and the distribution of dyad-score groups within each collaborative mode.

Research Question 2

3.9 ICAP Learning Engagement

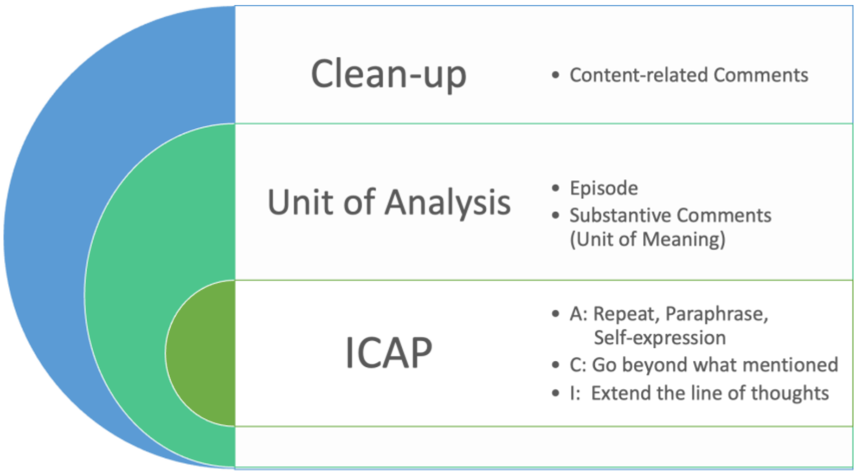
This study utilizes quantitative content analysis (QCA) on voice and text discussion to measure learning engagement. QCA is a research method used in the social sciences and other fields to systematically analyze and quantify the content of textual, visual, or audio materials. Through a structured, systematic code and code approach, the primary objective of quantitative content analysis is to organize and elicit meaning from the data collected and to draw realistic conclusions by identifying patterns, relationships, and trends within the content of materials (Rourke & Anderson, 2004). QCA is widely used in both audio and text discussion data to manifest meanings, intentions, communication dynamics, and group interactions (De Wever et al., 2006; Henri, 1992; Hou et al., 2008).

The ICAP hypothesis posits that the quality of learning is contingent upon the overt engagement of a pair of students with learning materials and with one another (Chi et al., 2008). In accordance with ICAP, if the collaborative modes produce different learning outcomes, there should be discernible differences in learning engagement within the collaborative modes, and vice versa. The ICAP hypothesis suggests that superior learning outcomes arise from heightened levels of learning engagement or increased learning engagement.

Thus, the learning engagement coding book is developed based on ICAP framework: Active (A), Constructive (C) and Interactive (I) to examine differences in learning engagement between dyads in synchronous voice-based and asynchronous text-based collaborative watching. With this notion, this study involves several steps to measure active, constructive and interactive learning (Chi et al., 2017) (See Figure 8).

Figure 8

Quantitative Content Analysis Process



1. Data Cleaning: The discussion corpus is refined through the selection of comments pertaining to the course and containing relevant ideas related to the concepts being taught.
2. Unit of Analysis Definition: The author establishes two levels of analysis, namely (a) an episode and (b) substantive comments. An episode is coded when content-relevant utterances revolve around the same topic and line of thought. Within each episode, substantive comments are segmented based on complete ideas or units of meaning, which can be words, phrases, sentences, or multiple sentences. Examples of substantive comments include:
    a. “Designer handbags” (word)
    b. “Because he didn't feel taken advantage of that.” (phase)
    c. “Anchoring is an arbitrary number based on being misinterpreted” (sentence),

   d. "Well, I don't want to say it is less serious issues, but I think, with Alief in general though it can be bad in some cases, I just think acknowledging that people have different beliefs, it allows us as individuals, to be more open minded yeah" (multiple sentences).

3. Coding of Learning Engagement: substantive comments are further coded into Active (A), Constructive (C) and Interactive (I) to assess the levels of learning engagement.
   a. Comments are coded as Active (A) when participants reiterate or rephrase what the video narrator or their partner has said. Self-expression that does not directly contribute to understanding or knowledge construction is also coded as Active, such as "I definitely buy into the idea of loss aversion with money a lot. I hate taking chances of losing money, so I always avoid investments lol".
   b. Comments that delve into ideas beyond what was covered in the instructional video or by their partner are coded as Constructive (C), for example, "You should end both good and bad experiences with a good experience at the end since duration does not play a role in determining a good and bad experience".
   c. Constructive comments are coded as Interactive (I) if they expand upon their peer's idea, such as "You provided a great example for this effect! However, I do think ending a bad experience at the peak can be more harmful than ending it on a good note at the end."

An independent samples t-test was conducted to compare the mean differences of substantive, active, constructive, and interactive comments between dyads the collaborative modes. The t-test is a widely used inferential statistical test that assesses whether there are significant differences in means between two independent groups. In this study, the independent groups were defined by the collaborative modes, namely the asynchronous and synchronous modes. By conducting an independent samples t-test, the author aimed to determine whether there were statistically significant disparities in the average frequencies of substantive, active, constructive, and interactive comments between the dyads in these two modes. This statistical analysis offers insights into the potential variations in the levels of engagement and interaction observed within the different collaborative modes. Additionally, descriptive statistics demonstrated the frequency and the distribution of learning engagement between and within the identified dyad-score groups to depict the overall dynamic of interaction in the collaborative modes.

Research Question 3

3.10 Discourses in Co-Construction Process

Based on social constructivism and ICAP, enhanced learning outcomes are observed when individuals actively participate in interactive moments, where they engage with others. As such, this study aims to closely analyze the interactive moments involving dyads, with the goal of gaining insights into their significance and impact on the learning process.

An interactive moment in this study refers to a situation in which a participant extends their partner's preceding line of thought, thereby establishing an "initial" and "response" structure in line with the Interactive (I) of the ICAP framework. To gain a deeper understanding of interactive moments, interactive discourse coding is employed to analyze the discourse used by dyads in co-constructing knowledge. Initially, the interactive discourse coding scheme adopted from Rakovic et al. (2020) consists of categories such as 'agree', 'disagree', 'give reason', 'request justification', 'ask question', 'build on', 'share', 'compare', 'make a claim', and 'answer'. Subsequently, the coding scheme is refined based on the collected data. The final interactive discourse coding book includes eight distinct discourses: explain, elaborate, struggle, hesitate, clarify, alternative, build on, and challenge. The coding book provides definitions and examples of each interactive discourse (See Table 2).

Table 2

Interactive Discourse Coding Book

Interactive Discourse	Operational Definition	Example
Explain	Define and explain definition of target concepts	"Loss aversion is basically trying to make sure you don't lose anything but instead either stay stagnant or gain something."
Elaborate	Elaborate and describe their understanding using such as examples, comparison, and analogy	"The example may be when you go to a mass casualty incident. You're going to treat the ones that are more injured, because you think, like they're the ones that need saving, but in reality, you're actually saving the most people treating the less injured people."
Struggle	Struggle to describe their understanding due to insufficient knowledge	"I don't know it's really hard to explain with gambling and investment because I don't know how investment works so it's sort of hard to explain."
Hesitate	Hesitate or delay to articulate their understanding.	"What was the example that they gave at the beginning with the movies, I didn't actually… I just rewind and watch it again."
Clarify	Ask to clarify, repeat, or re-explain what their partner said	"Oh yeah wait..what was that one again?" "Can you repeat what you just said again?"
Alternative	Propose alternative views	"You provided a great example for this effect! However, I do think ending a bad experience at the peak can be more harmful than ending it on a good note at the end." "um I think it just depends on what the peak and then the end effect. Maybe not so much duration. It just kind of depends, but I

		just think like it is about where the peak and the end where those things are”
Build on	Build on their partner’s idea to generate new line of thought	“Yeah, I agree my parents actually did the wine example to me. They gave me a cheap bottle and an expensive bottle but didn't tell me which was. But I prefer the bottle that was actually cheaper um.”
Challenge	Ask questions beyond video content to challenge oneself or their partner	“It seems that mental accounting much relates to money. I wonder what context this bias can also explain?” “But just wait, what if the event is like all peaks?”

3.11 Discourses Patterns in Co-Construction Process

This study analyzes the sequential probability of adjacent comments to investigate the discourse patterns that participants tend to construct in response to specific initial introductions. The likelihood of response discourse following a particular initial discourse is determined using n-gram analysis.

The term 'n-gram' refers to a contiguous sequence of n items, which can be words, characters, or other units of text, extracted from a given text or speech data. N-grams are used to identify patterns, relationships, and linguistic structures within the text. This technique involves extracting contiguous sequences of n items from a sample of text or speech (Ravi & Kim, 2007).

The Maximum Likelihood Estimation (MLE) technique is utilized to estimate the probabilities of observing specific n-grams in a text corpus. It calculates the likelihood of an n-gram occurring based on the frequency of its occurrence relative to the total number of n-grams in the corpus. Consequently, the resulting probabilities are constrained within the range of 0 to 1 (Jurafsky & Martin, 2023).

A bigram model is utilized in the study to calculate the probability of interactive discourses. A bigram is a specific case of an n-gram that calculates the probability of two items occurring together. The probability of interactive discourses corresponds to the likelihood of response discourse following a given initial discourse, where the structure follows an "initial → response" pattern. The computation of the bigram probability involves dividing the frequency of the specific interactive discourse of interest in the corpus by the total occurrence of the initial discourse within the same corpus.

$$P(w_{n-1}) = \frac{Count(w_{n-1}, w_n)}{Count(w_{n-1})}$$

For instance, the calculation of the sequential probability pertaining to the interactive discourse 'build on,' given the initial discourse 'elaborate' (i.e., 'elaborate' → 'build on'), entails dividing the frequency of occurrences in which 'elaborate' elicits 'build on' (52 instances) by the total number of occurrences where 'elaborate' serves as the initial discourse (143 instances). This computation yields a sequential probability of 0.36 (52/143). Consequently, when a participant commences a discussion with an elaboration, it becomes evident that roughly one third of the time their interlocutor is inclined to respond by expanding upon the initial discourse, further developing ideas.

3.12 Knowledge Co-Construction Phases

After identifying discourse and discourse patterns, our study employed thematic narrative analysis to categorize these patterns and identify co-construction phases across different dyad-score groups within collaborative modes. Thematic narrative analysis, a qualitative research approach rooted in thematic and narrative analysis, focuses on categorizing events and describing content or genre from the broader context (Riessman, 2008). Distinguishing between two closely related qualitative analysis approaches based on their purposes, as noted by McAllum et al. (2019), thematic analysis seeks to reveal overarching themes and meanings across cases through decontextualization of data. In contrast, narrative analysis aims to pinpoint pivotal moments in the development of events, facilitated by its process-oriented approach.

In our study, thematic narrative analysis was applied within the context of group interaction analysis to conceptualize the group process and provide context and descriptions of collaborative interactions among online students, as demonstrated by Fu et al. (2016). The primary objective of this analysis in our current study was to shed light on the process of knowledge co-construction and the evolving dynamics of interactions among dyads with high-high, high-low, and low-low scores in two distinct collaborative modes. Our unit of analysis was the discourse patterns identified within each collaborative mode, which were grouped based on their functional similarities.

CHAPTER 4

RESULTS

Research Question 1

4.1 Learning Outcomes

A one-way between-group analysis of covariance (ANCOVA) was performed to compare the mean scores of post-tests in sum, retention, and application between participants in asynchronous text-based collaborative mode (AT) and synchronous voice-based collaborative mode (SV) within two videos (referred to as overall videos), as well as within each individual video. Descriptive statistics, including the percentage of scores, mean, and standard deviation, are presented in Table 3.

The overall trend suggests that participants in the synchronous voice-based collaborative mode (SV) outperformed those in the asynchronous text-based collaborative mode (AT) in terms of sum and application scores, both in the overall videos and each individual video. However, participants in the AT achieved higher retention scores in the overall videos and each individual video. The findings suggest a discernible trend wherein participants engaged in SV tend to exhibit better performance and a greater ability to apply acquired knowledge in novel contexts. In contrast, individuals engaged in AT tend to show a higher tendency to retain the learned information from the videos.

After excluding outlier scores, the sum scores, retention scores, and application scores were normalized to mitigate potential biases or variations stemming from disparities in tests and to ensure comparability and standardization across different variables and measurements. The analysis of covariance (ANCOVA) results indicated no statistically significant differences in the overall videos between collaborative modes for sum scores [$F(1, 70) = 1.66$, $p = .20$] and retention scores [$F(1, 71) = .44$, $p = .51$]. However, a significant difference was found between participants in the SV and AT, with the SV mode yielding higher scores, $F(1, 71) = 4.64$, $p < 0.05$, $\eta_p^2 = 0.06$. The Table 3 demonstrated the means and standard deviations of percentage scores across two collaborative modes.

Table 3

Means and Standard Deviations of Scores

Scores	Collaborative Mode	Overall		Pricing Bias		Perception Bias	
		M	SD	M	SD	M	SD
Sum	Syn Voice	69.81	14.99	61.11	11.02	78.52	13.44
	Asyn Text	65.74	15.13	59.65	15.79	72.54	11.28
Retention	Syn Voice	78.70	17.18	74.99	20.01	82.40	13.37
	Asyn Text	80.55	14.63	78.07	14.75	83.33	14.43
Application	Syn Voice	63.27*	19.11	50.67	13.32	75.93*	15.36
	Asyn Text	55.86	19.96	47.37	20.90	65.36	14.10

Note. * p < .05

A one-way ANCOVA analysis was conducted to examine whether there were statistically significant differences in score means between the collaborative modes within each individual video: the pricing-bias video and the perception-bias video. In the pricing-bias video, the analysis revealed no significant differences across collaborative modes in terms of sum scores [$F(1, 35) = .13$, $p = .71$], retention scores [$F(1, 36) = .68$, $p = .42$] and application scores [$F(1, 35) = .46$, $p = .50$]. Similarly, in the perception-bias video, no significant differences were found in sum scores [*F* (1, 33) = 1.35, *p* = .25] and retention scores [*F* (1, 33) = .18, *p* = .67] in perception-bias video. However, participants in SV outperformed AT in application scores [*F* (1, 34) =, *p* < .05, η_p^2 = .13].

4.2 Dyad-Score Groups

Since a significant difference was detected in application scores between synchronous voice-based and asynchronous text-based collaborative modes, K-means cluster analysis was applied to these scores. This analysis aimed to investigate the connection between collaborative modes and dyad performance by shifting the focus from individual scores to dyad scores. The analysis allowed for a comprehensive comparison and examination of how collaborative modes relate to the performance of dyads as a collective unit.

K-means is a widely employed clustering algorithm utilized to partition a dataset into K distinct clusters. It iteratively assigns data points to clusters based on their proximity to cluster centroids with the aim of minimizing within-cluster variance. Consequently, four final cluster centers were identified and classified, as illustrated in Figure 9. A scatter plot was employed to visualize the distribution of members within each cluster (See Figure 10),

Cluster 1 consisted of dyads whose scores were centered around the mean for both individuals, resulting in a z-score of 0 or falling within one standard deviation of the mean. This cluster, referred to as the "Average Cluster," represented dyads characterized by relatively balanced and consistent performance around the mean.

Cluster 2 comprised dyads in which both members achieved high scores, forming the "High-High Cluster." These dyads demonstrated consistently high-performance levels.

Cluster 3 encompassed dyads in which both members obtained low scores, constituting the "Low-Low Cluster." These dyads exhibited consistently low-performance levels.

Cluster 4, also known as the "Low-High Cluster," represented dyads with notable score discrepancies, where the performance differences between members were approximately 2 standard deviations or more. For instance, within the same dyad, one member displayed scores lower (z-score = -1.64), while the other member exhibited higher scores (z-score = 1.14), highlighting a considerable imbalance in performance.

Figure 9

The Clusters of Dyad-score Groups

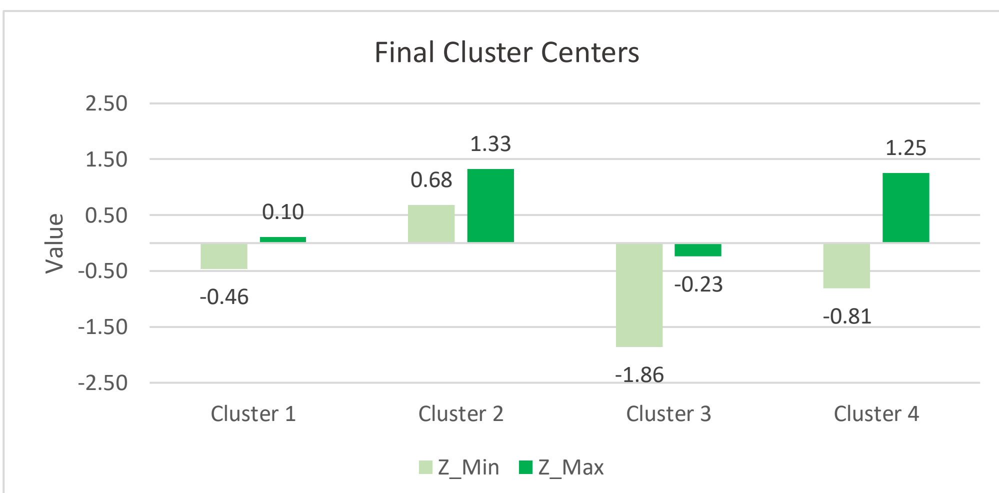
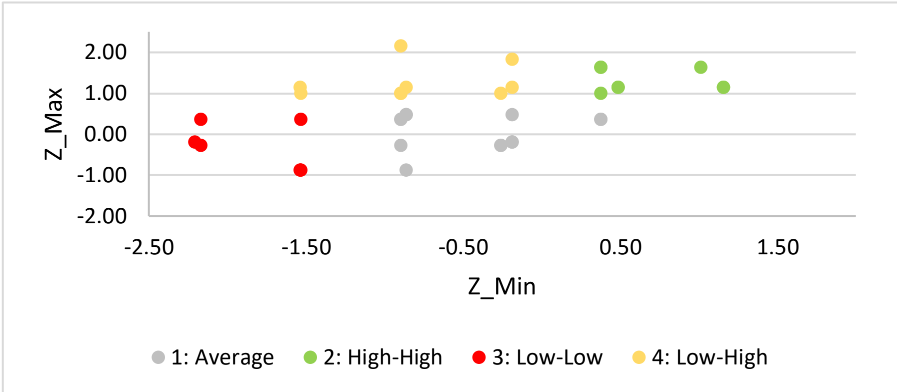


Figure 10

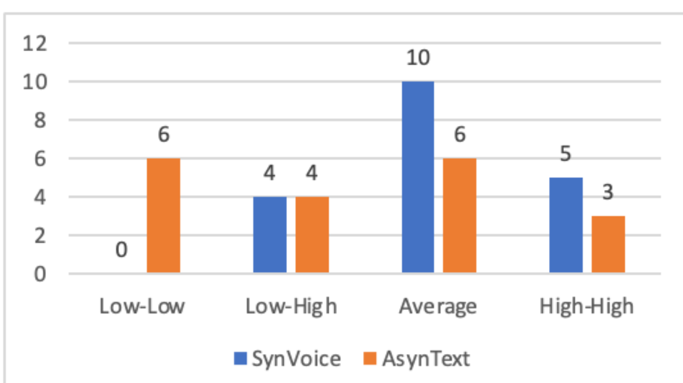
Scatter Plot of Dyad-score Groups



The analysis of dyad performance across different collaborative modes yielded noteworthy findings, particularly when examining the representation of SV and AT dyads within each cluster. The SV demonstrates a higher representation in the Average and High-High Clusters (n =15), suggesting a potential advantage in achieving balanced and high-performance levels. On the other hand, the AT exhibits a greater presence in the Average and Low-Low Cluster (n =12), suggesting a tendency for consistently balanced and lower performance. However, dyads from both collaborative modes equally represent the mixed performance group, including the Low-High Cluster (See Figure 11).

Figure 11

The Number of Dyads in the Colloborative Modes by Cluster



Research Question 2

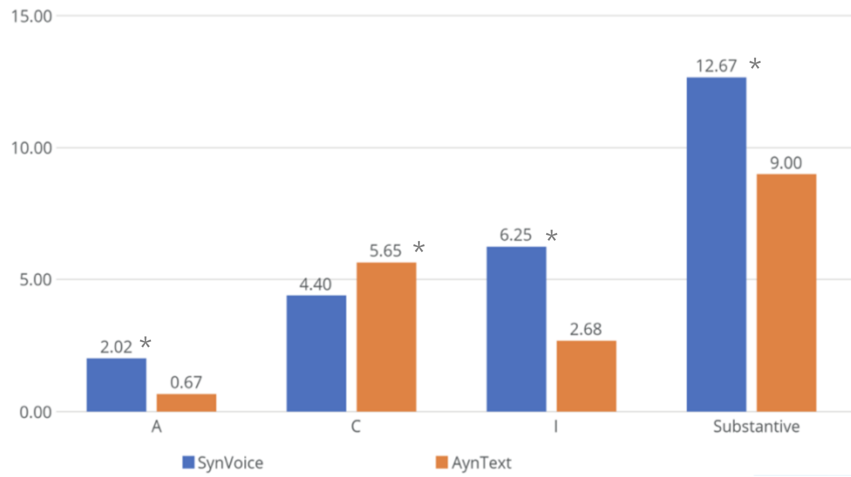
4.3 ICAP Learning Engagement

Learning Engagement in Collaborative Modes. The overall trend revealed that dyads in the SV were more likely to engage in discussions, as evidenced by higher numbers of substantive comments. Specifically, dyads in the SV generated more active comments (i.e., A in ICAP), indicating involvement and responsiveness to the learned concepts, such as repeating what was said in the video or by their partners. However, it is important to note that Active comments do not directly contribute to the generative process of constructing or revising mental representations. Furthermore, dyads in the SV also produced more interactive comments (i.e., I in ICAP), indicating interaction in dialogue where one participant adopts and builds upon another participant's ideas to co-construct knowledge. This interactive nature of the dialogue suggests a collaborative and mutually supportive learning process within the SV.

In contrast, participants in the AT generated more constructive comments (i.e., C in ICAP), indicating a higher occurrence of generative monologues in the co-construction process. These comments involved participants sharing their own thoughts, perspectives, and ideas without direct interaction with others. Similar to the effects of explaining concepts to oneself or to a fictional character, constructive comments allowed participants to articulate and externalize their understanding, enhancing comprehension and deepening the construction of mental representations (Chi & Wylie, 2014). Below figures presents the frequencies of Active (A), Constructive (C), Interactive (I), and substantive comments in both the pricing-bias and perception-bias videos (See Figure 12).

Figure 12

Learning Engagement in the Collaborative Modes



An independent t-test was conducted to examine the engagements generated by participants in AT and SV. The results revealed that participants in the AT generated significantly more constructive comments [$t(55) = 2.46$, $p < .05$, $d = .46$]. On the other hand, dyads in the SV created more active comments [$t(55) = 3.72$, $p < .001$, $d = .70$], interactive comments [$t(55) = 6.10$, $p < .001$, $d = 1.14$], and substantive comments [$t(55) = 4.45$, $p < .001$, $d = .83$]. Table 4 shows the means and standard deviation of substantive comments and comments associated to active, constructive, and interactive learning engagement.

Table 4

Means and Standard Deviations of Learning Engagement

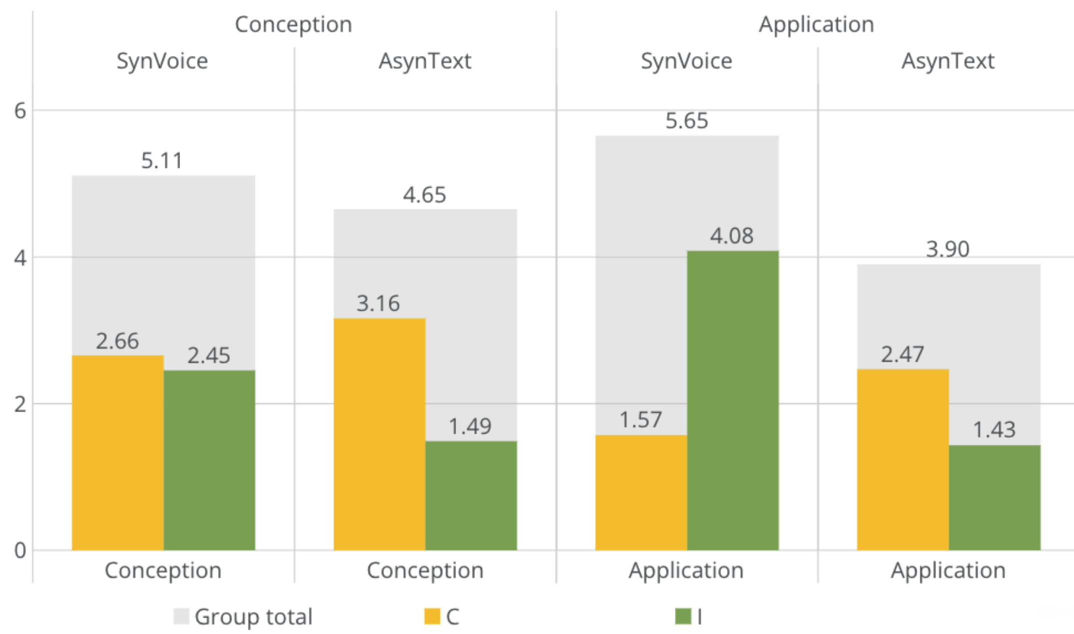
Learning engagement	Collaborative Mode	N	M	SD
Active	SynVoice	57	2.02**	2.15
	AsynText	57	0.67	1.69
Constructive	SynVoice	57	4.40	2.99
	AsynText	57	5.65*	2.37
Interactive	SynVoice	57	6.25**	3.65
	AsynText	57	2.68	2.45
Substantive	SynVoice	57	12.67**	5.51
	AsynText	57	9.00	2.85

Note. * $p < .05$; ** $p < .001$

Learning Engagement by Prompts. When examining generative comments, which encompass both constructive and interactive comments during the conception and application prompts, SV dyads prove to be more generative in both conception and application compared to AT dyads. However, the difference in the generative comments related to conception between SV and AT dyads is narrower compared to the generative comments related to application. This suggests that dyads in both collaborative modes are similarly engaged in conception, but they become more distinct in the application prompts. Due to the relatively more static nature of conception, dyads in both collaborative modes described the concepts with more constructive comments. However, SV dyads were more interactive in more dynamic prompts such as application. In general, SV dyads tend to prioritize application prompts with more interactive comments, whereas AT dyads tend to focus more on the conceptual aspects with more constructive comments. Figure 13 demonstrated the average number of constructive and interactive comments in each concept, based on conception and application prompts.

Figure 13

Learning Engagement by Prompts



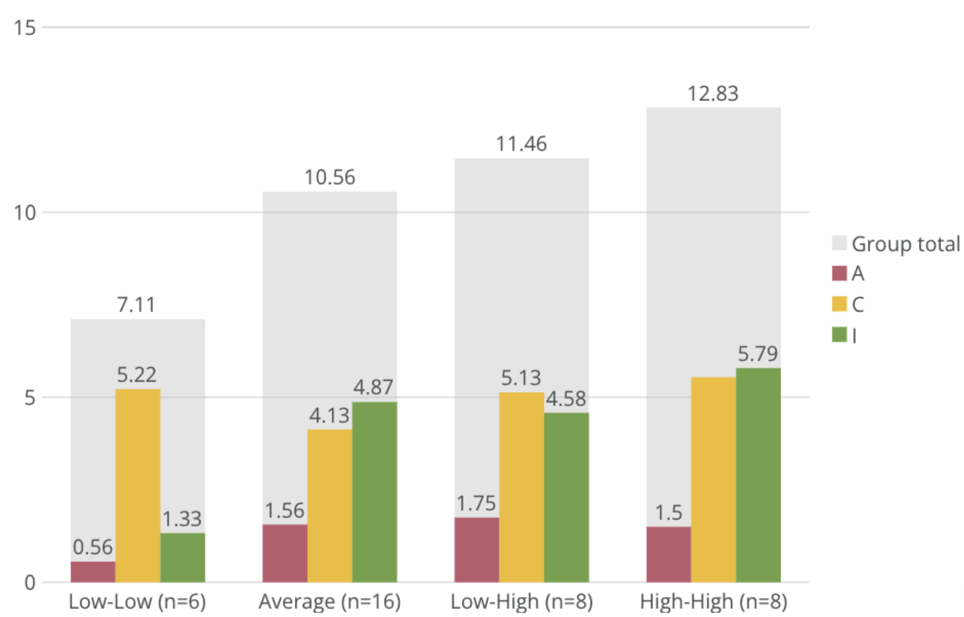
4.4 ICAP Learning Engagement of Dyad-Score Groups

Learning Engagement in Dyad-Score Groups. Further analysis of the dyad score clusters revealed discernible patterns in learning engagement. While the distribution of learning engagement in the average, low-high, and high-high clusters exhibited similarities, the low-low cluster showed notable distinctions with considerably low levels of substantive comments, particularly in terms of interactive engagement. The least substantive comments in a low-low cluster indicated a passive disposition, making this cluster the most inert group among the four clusters. In contrast, the high-high cluster demonstrated the highest engagement, as reflected by the most substantial comments.

Both the average and high-high clusters displayed a greater prevalence of interactive comments compared to constructive comments, with the high-high cluster having the highest frequency of interactive comments. On the other hand, the low-high and low-low clusters exhibited a higher occurrence of constructive comments in comparison to interactive comments (See Figure 14).

Figure 14

Overall Learning Engagement

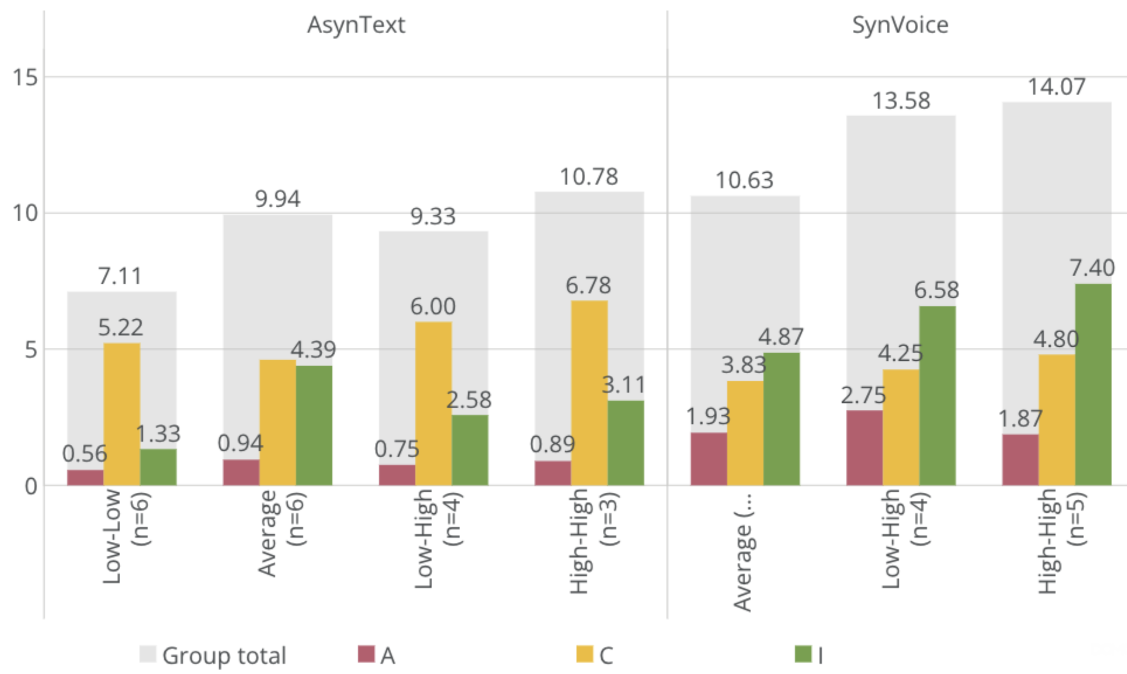


When analyzing the distribution of learning engagement within each cluster, grouped by collaborative modes, a consistent pattern emerges. In the SV, dyads consistently exhibit a higher frequency of interactive comments across all clusters, while the AT leans towards more constructive comments. Notably, the high-high cluster in the SV demonstrates the highest level of engagement with interactive comments, whereas the high-high cluster in the AT displays a stronger focus on constructive comments. The low-

low, average, and low-high clusters follow the same ordered pattern within their respective modes (See Figure 15).

Figure 15

Learning Engagement in Collaborative Modes



The number of constructive and interactive comments in SV is relatively closer compared to the AT. This suggests a greater balance between individual and dyad contributions in the SV mode. However, the low-high clusters in both modes exhibit the greatest disparity between the numbers of constructive and interactive comments, indicating a tendency for one partner to dominate the discussion within a dyad.

Learning Engagement in Posts and Turns. An AT dyad generated an average of 3.20 posts per concept, with a majority of their responses directed towards conception prompts (2.22 posts) and fewer towards application prompts (1.76 posts). In contrast, among an average of 5.39 turns per concept created by a SV dyad, a larger proportion was focused on addressing application prompts (3.06 turns) as opposed to conception prompts (2.33 turns).

Regardless of whether the responses are posts in AT or turns in SV, all dyad-score groups in the collaborative modes, except for low-low AT, exhibit a consistent pattern with each turn and post containing the similar number of generative comments within the conception prompts, ranging from 4.44 to 5.83 comments. However, in application prompts, dyads showed different generative amount either within or between the collaborative modes.

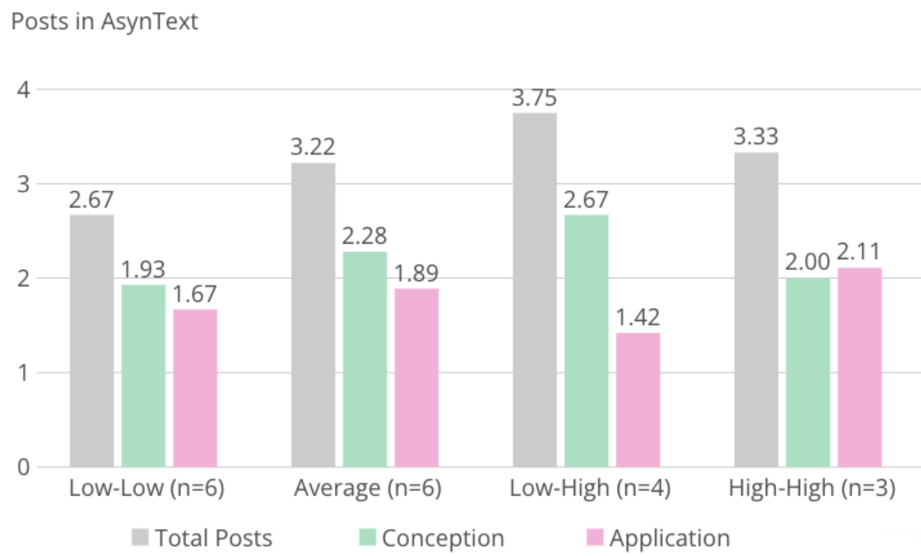
High-high SV dyads exhibit a distinct contrast in their turns and generative comments. They produce the fewest number of turns (4.8 turns), yet each turn includes the highest number of generative comments (6.47 comments), especially within interactive comments in application prompts (7.33 comments). This indicates that each dyad makes substantial and well-developed contributions in each concept, in contrast to other dyad-score groups in SV that tend to be more trivia and fragmented. A similar trend is also observed in high-high AT, where they consistently have the most generative comments in each post (6.18 comments).

In the conception prompts, both collaborative modes consistently show that low-high groups have the highest number of posts (2.67 posts) and turns (2.75 turns), as well as the most generative comments (5.58 and 5.83 comments, respectively). However, when it comes to application prompts, especially in low-high AT (3.00 comments), the number of generative comments decreases. This suggests that low-high groups put in effort to gain conceptual understanding but may face challenges when applying it to application prompts.

Conversely, in the application prompts, high-high groups in both collaborative modes consistently exhibit the highest number of generative comments (5.45 comments in AT, 7.33 comments in SV). This aligns with the findings in between-mode prompt analysis (See Figure 13), indicating that SV dyads tend to be more generative and interactive in application prompts. The score-dyad prompt analysis further indicates that high-high dyads consistently invest more effort in application prompts, regardless of the collaborative mode. Figure 16 shows the number of posts per concept of a dyad in each AT dyad-score group and the number of generative comments per post based on conception and application prompts. Similarly, figure 17 shows the number of turns per concept of a dyad in each SV dyad-score group and the number of generative comments per turn based on conception and application prompts.

Figure 16

The Number of Posts and Learning Engagement (AT)



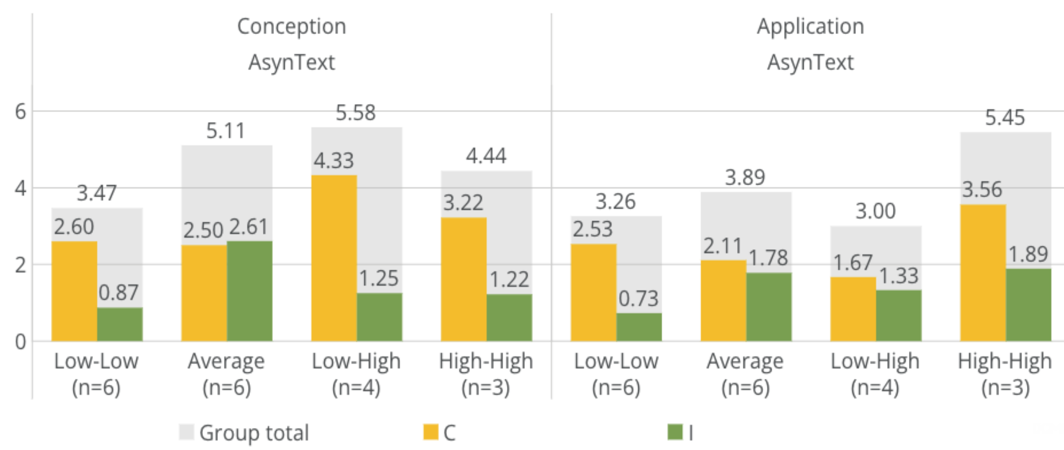
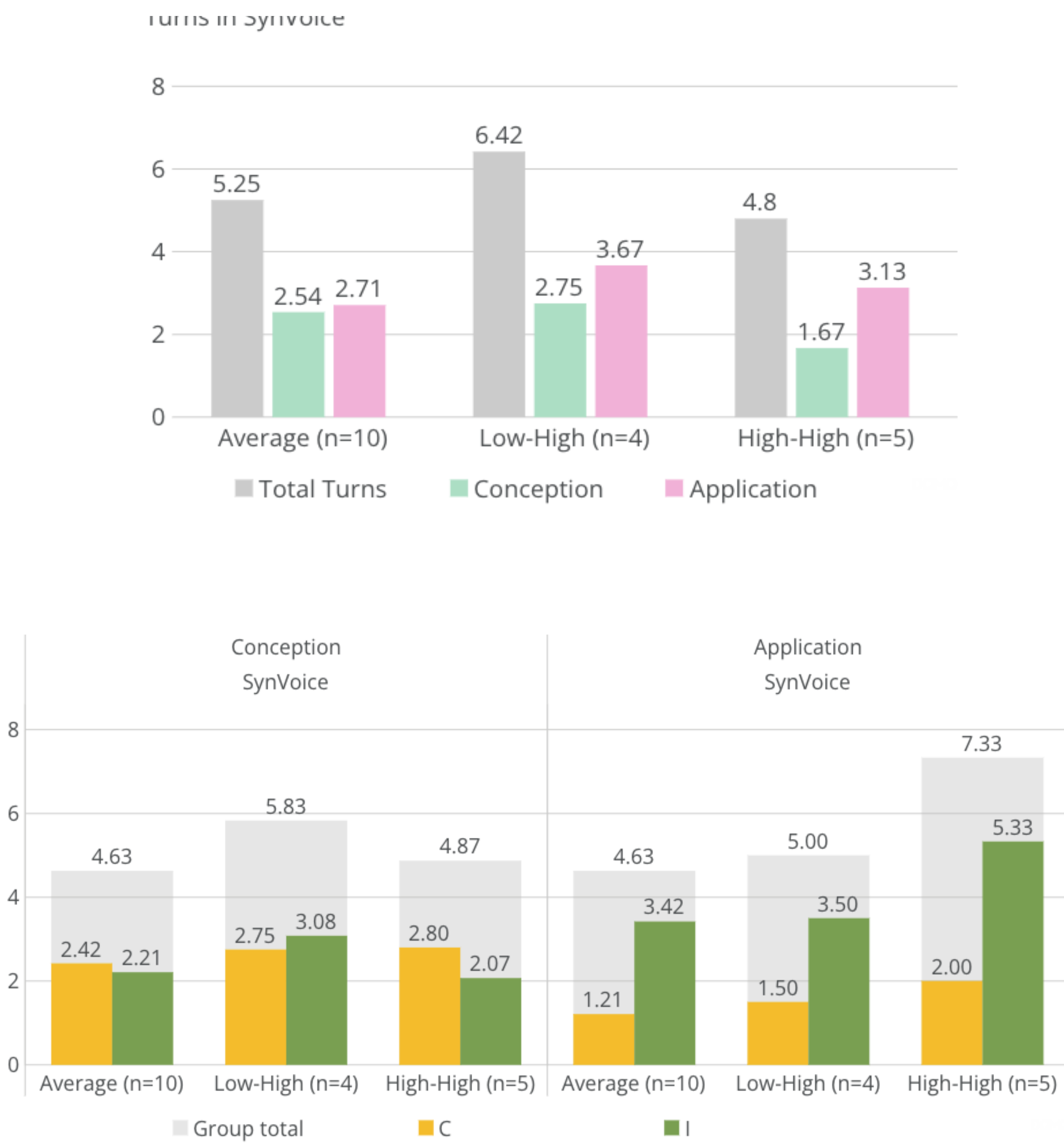


Figure 17

The Number of Turns and Learning Engagement (SV)



Research Question 3

4.5 Discourses in Co-Construction Process

The findings revealed that dyads in both collaborative modes employed a variety of interactive discourses during the co-construction process. Overall, those in the synchronous voice-based mode exhibited higher frequency and greater diversity of discourse in both the initial and response stages (See Table 5).

The number of initial and response discourse in collaborative modes

Table 5

Discourse Frequency in Collaborative Modes

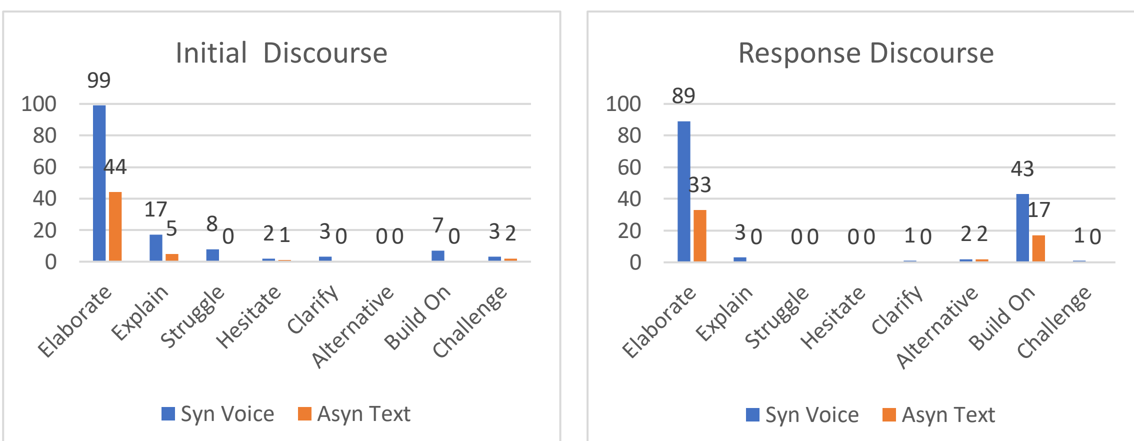
Interactive Discourse	Synchronous Voice-based Mode		Asynchronous Text-based Mode	
	Initial	Response	Initial	Response
Elaborate	99	89	44	33
Explain	17	3	5	0
Struggle	8	0	0	0
Hesitate	2	0	1	0
Clarify	3	1	0	0
Alternative	0	2	0	2
Build On	7	43	0	17
Challenge	3	1	2	0
Total Interactive Discourse	139	139	52	52
Total Substantive Comments	708		516	
Percentage of Interactive Discourse	19.63%		10.08%	

Among the 191 instances of interaction or interactive moments, where the initial comment elicits a response (i.e., initial → response), dyads in both collaborative modes predominantly used 'elaborate' and 'explain' to initiate discussion. However, the SV dyads also employed additional discourses such as 'struggle', 'build on', 'challenge', 'clarify', and 'hesitate' as part of their initiation. In contrast, AT dyads tended to respond to the initial discourses that were more explicit and demanding in nature, such as 'challenge' and 'hesitate'.

Similarly, in their responses, dyads in both collaborative modes primarily utilized 'elaborate' and 'build on' to enrich the ideas. They also occasionally employed 'alternative' to propose other possible explanations. In the fast-paced and sometimes unorganized environment of the synchronous voice-based mode, dyads further incorporated response discourses aimed at resolving ambiguity in communication, such as 'explain' and 'clarify'. Furthermore, due to the dynamic dialogue style, SV dyads occasionally deviated from directly responding to the prompt or mutually elaborating with their partner. Instead, they questioned the claim presented in the video or by their partner through discourses like 'challenge' (See Figure 18).

Figure 18

Discourse Frequency in Initial and Response

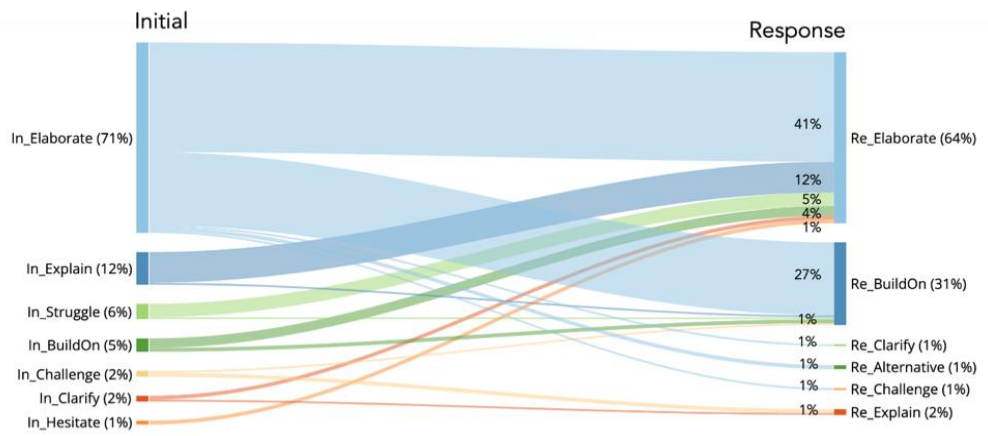


4.6 Discourses Patterns in Co-Construction Process

The SV exhibited a greater variety of interaction patterns compared to the AT. Among the 139 interactive moments in SV discussions, sixteen unique interaction patterns were identified. The most common interactions included 'elaborate' → 'elaborate' (41%), 'elaborate' → 'build on' (27%), and 'explain' → 'elaborate' (27%). Other less common interactions included 'explain' → 'elaborate' (12%), 'struggle' → 'elaborate' (5%), and 'build on' → 'elaborate' (4%). The remaining patterns, such as 'elaborate' → 'clarify' (0.7%), 'elaborate' → 'alternative' (1.4%), 'elaborate' → 'challenge' (0.7%), 'explain' → 'build on' (0.7%), 'struggle' → 'build on' (0.7%), 'challenge' → 'build on' (0.7%), 'clarify' → 'elaborate' (1.4%), 'clarify' → 'explain' (0.7%), and 'hesitate' → 'elaborate' (1.4%), were less frequently observed (See Figure 19).

Figure 19

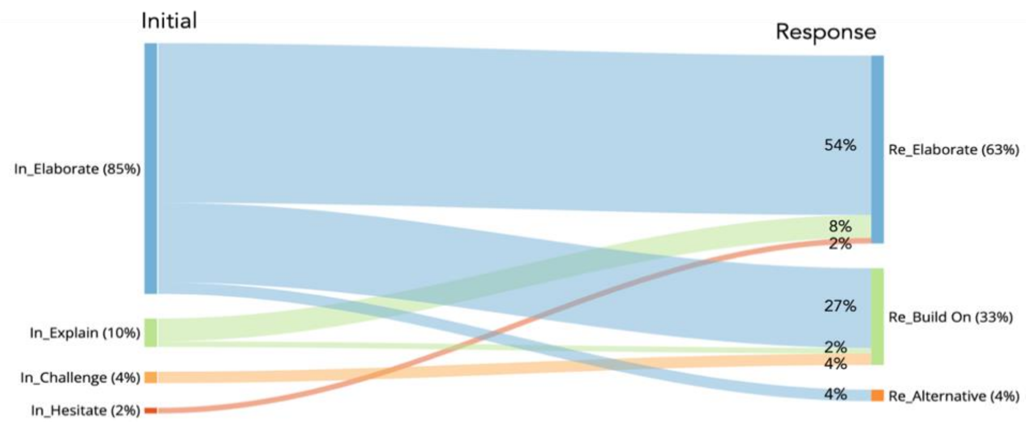
Discourse Patterns in SV



In the AT, among the 52 interactive moments, seven unique discourse patterns were identified. Similar to the SV discussions, the most common patterns were 'elaborate' → 'elaborate' (54%) and 'elaborate' → 'build on' (27%). Other patterns included 'explain' → 'elaborate' (7.7%), 'elaborate' → 'alternative' (3.9%), 'explain' → 'build on' (1.9%), 'challenge' → 'build on' (3.9%), and 'hesitate' → 'elaborate' (1.9%) (See Figure 20).

Figure 20

Discourse Patterns in AT



There were notable discrepancies in the interaction patterns between the two collaborative modes. In the AT, when a participant provided an elaboration, no participant responded with 'clarify' or 'challenge' (e.g., 'elaborate' → 'clarify'). The absence of 'clarify' discourse suggests that the comments in this mode were well-organized and easy to understand, eliminating the need for clarification. It is also possible that some participants were passive and did not actively engage in the discussion, resulting in the lack of 'challenge' after an elaboration.

However, when a participant explicitly challenged their partner, it triggered the 'build on' discourse to extend the line of thoughts, unlike the challenge in the SV which elicited either 'build on' or 'explain' responses. Additionally, several discourses were missing in the initial stage, such as 'struggle', 'clarify', and 'build on', indicating that either no participant used these discourses to initiate the discussion, or no participant responded to any of these initial discourses. Consequently, no interactive moments associated with these initial discourses were found, and the occurrence of interaction patterns was not identified.

The sequential probability of the adjacent comments is analyzed to examine discourse patterns that dyads are likely to employ in response to specific initial comments. In the SV, when dyads initiate the discussion with an elaboration, more than half of the time, their partner is likely to 'elaborate' to continue expanding on the original line of thought ($p = 0.58$), and there is a fair chance of 'build on', creating a new line of thought ($p = 0.38$). However, if the initial elaboration is unclear or poorly argued, there is a possibility that their dyad will respond with 'clarify' ($p = 0.01$), 'alternative' ($p = 0.02$), or 'challenge' ($p = 0.01$) (See Figure 4). Similarly, participants in the AT are also likely to respond to their partner's elaboration with 'elaborate' ($p = 0.64$) and 'build on' ($p = 0.32$). Surprisingly, irrespective of the quality of the initial elaboration, participants in this mode display a tendency to neither request clarification nor challenge their partner to provide more detailed explanations. However, they do occasionally present alternative ideas ($p = 0.05$) as a courteous way to express disagreement (See Figure 21).

When dyads initiate the act of 'explain' in both collaborative modes, there is a high probability that the other participant will 'elaborate' and enhance the previous ideas ($p = 0.94$ in SV and $p = 0.80$ in AT). Furthermore, there is a fair chance that the initial explanation will inspire new lines of thought for the other participant to 'build on' ($p = 0.13$ SV and $p = 0.20$ in AT mode).

In the SV, when dyads encounter a challenge, they typically respond by providing an explanation ($p = 0.67$) or building on the challenge ($p = 0.33$), while in the AT, challenges often lead to a significant generation of new ideas and perspectives ($p = 1.00$). Additionally, when participants experience hesitation during the discussion, they tend to counteract it by offering more extensive elaboration ($p = 1.00$) in both modes, aiming to support and encourage their partner to continue the discussion.

However, only dyads in the SV initiate 'struggle', 'clarify', and 'build on' and create their associated interaction patterns. When one participant struggles to explain or elaborate, there is a high probability ($p = 0.88$) that their partner will respond with an 'elaborate' discourse, providing detailed support and assistance. Additionally, there is a lower probability ($p = 0.12$) of the subsequent discourse being 'build on', indicating an attempt to shift the focus to simpler aspects of the previous line of thought.

In situations where uncertainty or the need for clarification arises, participants tend to prioritize 'elaborate' ($p = 0.67$) to further develop their ideas or 'explain' ($p = 0.33$) the previous ideas again, emphasizing the importance of clarity and understanding within the discussion.

SV dyads frequently utilize the 'build on' discourse as a response to the preceding idea. However, in situations where other subsequent discourses are present, the 'build on' response becomes an initial discourse. The discourse pattern typically follows a sequence of "initial comment → 'build on' (initial/response) → response comment", which indicates the flow of the conversation. The subsequent discourses that follow 'build on' are primarily 'elaborate' ($p = 0.71$) and 'build on' ($p = 0.29$), highlighting the continuity and dynamic nature of the discussion as dyads actively build upon and expand upon the evolving ideas

The identified interaction patterns suggest that both SV and AT dyads commonly use the 'elaborate' and 'explain' discourses to initiate discussions and gradually co-construct knowledge through the use of 'elaborate' and 'build on' discourses. The prevalence of 'elaborate' indicates a desire to provide further detail and develop the previous line of thought, while the occurrence of 'build on' reinforces the collaborative nature of the exchange as participants contribute their own insights and perspectives.

While there is no direct evidence of struggle or the need for clarification in the AT dyads, dyads in the SV perceive struggles or ambiguity as opportunities to address misalignments and work towards co-constructing knowledge and establishing a shared understanding (See Figure 20).

In the AT, although dyads may respond with 'build on', they often disregard the continuity of the new line of thought (e.g., initial comment → 'build on'). Conversely, in the SV, dyads keep continuing expand and extent the the new line of thoughts with more detials or another new line of thoughts (e.g., initial comment → 'build on' (initial/response) → 'elaborate' or 'build on'), indicating a tendency to maintain the progression of ideas (see Figure 22).

Figure 21

The Sequential Probability of Discourse Pattern in SV

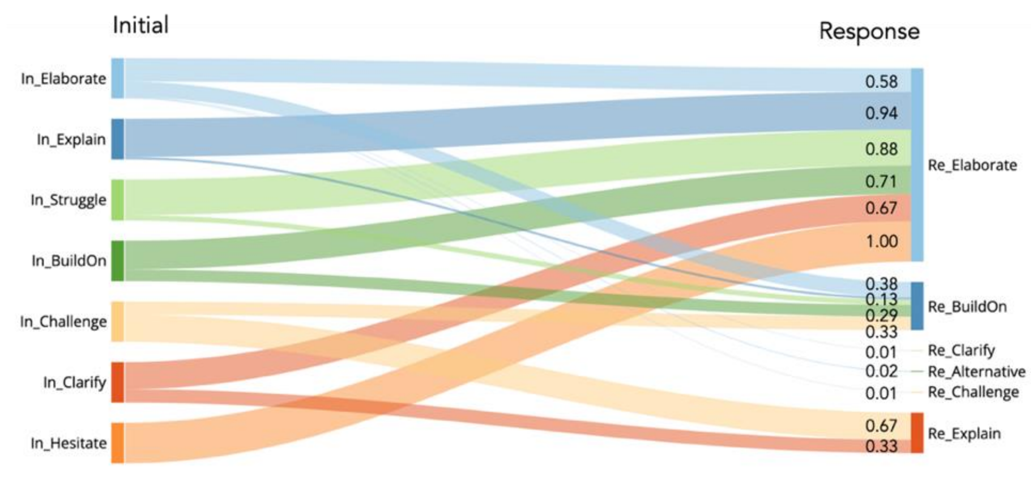
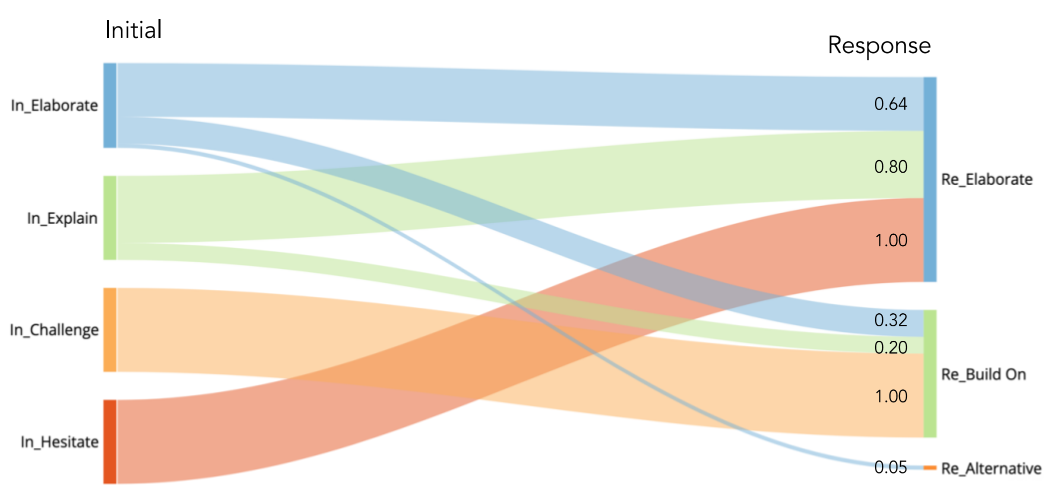


Figure 22

The Sequential Probability of Discourse Pattern in AT



4.7 Knowledge Co-Construction Phases

Based on discourse and discourse patterns, this study identified knowledge co-construction phases to elucidate the process of knowledge co-construction and the evolving nature of interactions among dyads with high-high, high-low, and low-low scores in two distinct collaborative modes. Based on themeatic narrative analysis, the interactive discourses observed within group interactions can be classified into three phases: co-explanation, negotiation, and application, based on their functional similarities.

The co-construction phases offer insights into how interactive dynamics, including the number and ratio of constructive and interactive engagement, impact the overall quality of discussions among dyad-score groups and across collaborative modes. Moreover, on a micro level, within each phase, the results illustrate how dyads employ a wide array of interactive discourses to collectively build knowledge. These findings provide a comprehensive understanding of the specific strategies and approaches utilized by dyads in each collaborative mode throughout the knowledge co-construction process.

4.7.1 Co-Explanation Phase

Explain, Elaboration: Collaboratively explain and elaborate on ideas to construct the shared understanding

This phase is commonly observed in dyad interactions, where most dyads initiate discussions by presenting their definitions of the target concepts, aiming to share their initial understanding. They engage in explaining and conceptualizing the learned concepts to establish a broad understanding, primarily focusing on the "what is..." aspect, often overlooking specific details pertaining to instances, procedures, and applications.

Dyads collectively explored ambiguous concepts, leading to the revision and elaboration of incomplete explanations. This process involves a reciprocal movement between concretizing and conceptualizing to revise the entire representation. The co-construction endeavor of dyads goes beyond mere information accumulation as they gradually transform their understanding and build the collective knowledge through interactive exchanges (Scardamalia & Bereiter, 2014). Discourse patterns associated with this phase, ranked by frequency, include: 'elaborate' → 'elaborate', 'explain' → 'elaborate', and 'explain' → 'elaborate'.

4.7.2 Negotiation Phase

Struggle, Hesitate, Clarify, Alternative: Negotiate to reconcile misalignment and clarify misunderstandings to further co-construct knowledge

This phase is typically more prevalent in collaborative modes that involve voice-based interactions, as the social presence plays a vital role in facilitating interactive momentum and the need for negotiation. Dyads enter this phase when uncertainty arises, either due to fragmented comprehension or misalignment with their partner, prompting them to actively seek additional insights and perspectives through interactive engagement to revise and realign their understanding of the subject matter.

Uncertainty arises when dyads encounter challenges in describing learned concepts or have doubts regarding their comprehension and articulation of those concepts. This uncertainty sometimes leads them to hesitate and become passive, hindering their willingness to assert their opinions. The struggle in these interactions may not solely stem from individual comprehension limitations but can also result from communication ambiguity during discussions, necessitating the clarification of intended meanings. This process of clarification not only enhances communication and promotes better understanding for their partners, but also enhances one's own understanding as they engage in learning by explaining to others (Lee et al., 2022; Roscoe & Chi, 2008).

The negotiation phase also plays a critical role in creating a space for sharing alternative views and negotiating to realign misalignment. Given the peer-to-peer nature of the discussion, where participants lack expertise, the ability to verify the accuracy of their discourse is compromised. Consequently, it becomes essential to solicit and incorporate diverse perspectives, encompassing individual and expert insights obtained from learning videos, to enhance collaborative inferences (Deiglmayr & Spada, 2010; Hull & Saxon, 2009; Stahl, 2006). This inclusive and collective approach enables the evaluation of different explanations and contributes to a more comprehensive understanding of the subject matter.

Discourse patterns associated with this phase, ranked by frequency, include: 'struggle' → 'elaborate', 'elaborate' → 'clarify', 'elaborate' → 'alternative', 'elaborate' → 'challenge', 'struggle' → 'build on', 'challenge' → 'build on', 'clarify' → 'elaborate', 'clarify' → 'explain', and 'hesitate' → 'elaborate'.

4.7.3 Application Phase

Build On, Challenge: Go beyond requirements to seek opportunities to deeply engage in structural and application levels

During the application phase, dyads move beyond the foundational requirements and begin to delve deeper into the subject matter by actively engaging at both structural and application levels. The application phase demonstrates a higher level of cognitive presence as dyads integrate information and transition into a more focused and structured phase of sense-making (Garrison & Anderson, 2003). Their mental representation of concepts becomes more interconnected and less differentiated as their knowledge becomes encapsulated. Once complexity is reduced through the construction of a meaningful mental framework, the mapping between concept layers and across domains is achieved (Akyol & Garrison, 2011; Chi & Ohlsson, 2005; Schmidt & Boshuizen, 1993).

This phase typically occurs when dyads have achieved a basic understanding of the targeted concepts and established some level of shared understanding. As they progress, dyads start to build upon each other's existing knowledge, expanding upon concepts and ideas, while also challenging assumptions and exploring different perspectives. The increased connectedness in this phase fosters creativity and encourages participants to explore innovative solutions and novel perspectives beyond what they learn from videos (Akyol & Garrison, 2011). This process enables dyads to gain a more profound understanding of the topic and develop critical thinking skills, empowering them to apply their insights in practical and meaningful ways.

Discourse patterns associated with this phase, ranked by frequency, include: ‘elaborate' → 'build on' (27%), 'build on' → 'elaborate', 'explain' → 'build on', 'struggle' → 'build on', and 'challenge' → 'build on'.

4.8 Knowledge Co-Construction Phases in Dyad-Score Groups

4.8.1 Low-Low Group

Low-Low: Asynchronous Text-based Group (AT). Within the low-low group, consisting solely of AT dyads, interactions are scarce. Not only do they provide brief responses to prompts but also exhibit minimal engagement in reviewing their partner's replies or engaging in collaborative endeavors to deepen their understanding jointly. Whenever these dyads attempt co-construction, their sole interactive discourse pertains to elaboration (i.e., explain/elaborate → elaborate), often marked by brevity and triviality. No discernible signs of negotiation or higher-order cognitive presence are evident among dyadic interactions, reflecting a limited depth of meaningful exchange (Garrison et al., 2010).

The diminished social interaction can potentially lead to reduced motivation, resulting in certain dyads within the text-based mode adopting a more passive stance, wherein they treat the co-watching activity as an individual task. Consequently, the potential benefits of collaborative video watching are nullified, rendering the learning outcomes equivalent to those achieved through traditional individual video watching. In the given excerpt, each participant has briefly defined Alieve, yet they have not provided detailed examples or practical applications to support their explanations. For instance, participant 60 merely mentioned racial bias as an instance of bad Alieve, without delving into how and why racial bias can overshadow rational thinking and potentially jeopardize oneself or others. On the other hand, participant 70 engaged in the discussion but failed to assist participant 60 in exploring the concept of racial bias and its connection to Alieve. Instead, participant 70 introduced their own perspective on Alieve, specifically related the good Alieve to safety awareness. Although they did provide examples of both positive and negative applications of Alieve, further clarification and elaboration are necessary to justify and comprehend the implications of these applications.

Participant 60:	"An alieve is a belief grounded in bias and lack of reason. Racial bias is an example of a bad alieve."
Participant 71:	"An alieve is a sub conscious thought process based around bias and no true rational thinking. I think an alieve that is good and bad could be when you walk down a dark street at night. You believe it is safe but you may have an alieve that you will be attacked."

Notably, dyads within this group manifest the lowest level of generative commentary, encompassing both constructive and interactive components. The observed deficiency in generative comments and corresponding low scores corroborate previous research, which highlights the association between generative learning engagement and positive learning outcomes (Fiorella et al., 2020; Jacob et al., 2020; Muldner et al., 2014).

Since interactive contributions are limited, low-low dyads often only spend time in the co-construction phase but rarely progress to higher-level phases. Although their contributions are brief and seldom prompt further discussion, a few initiations and responses primarily focus on explaining and elaborating concepts. Thus, the predominant and exclusive discourse pattern found in this group is 'explanation/elaboration' followed by elaboration in an initiate-response style.

4.8.2 Low-High Group

In the high-low group, dyads consisted of participants from both AT and SV. Compare with low-low group, the AT in low-high group has more generative (i.e., constructive and interactive) comments. Within the same low-high group, the SV had more generative comments than the AT. This discrepancy between the modes in low-high group was typically observed during the negotiation phase and was often influenced by the higher-scoring partner.

Upon a closer examination of the generative comments, it was observed that while the ratio of constructive and interactive comments in SV dyads in the high-low group was similar to other groups, the difference between the comments in low-high AT was the largest among the groups. The imbalanced proportion of constructive comments compared to interactive comments suggests that one partner might have been more active in typing, while the other remained relatively passive. Due to the sole contribution of an active partner, constructive comments can considerably outweigh interactive comments, which may make them more susceptible to receiving different scores and being placed in the high-low group.

Low-High: Asynchronous Text-based Group (AT). In the AT, dyads encounter challenges such as "ghosting" or "thread dead," which transform collaborative video watching into an individualized activity and negate the benefits of collaborative learning. While "thread dead" remains a prevalent and noteworthy issue in asynchronous collaboration (Hew et al., 2010; Hewitt, 2005; Oztok et al., 2012), the severity of its impact varies depending on the sequence of contributions. The current study found that all first responders in the high-low group simply responded to the initial prompt and refrained from revisiting the discussion to engage with their partner. As predicted by ICAP (Interactive, Constructive, Active, and Passive) theory, these first responders primarily engage in constructive but not interactive interactions. In contrast, the second responders in the low-high group read their dyad's comments and subsequently interacted with the initial post. As expected, all second responders from the dyads in the text-based mode of the high-low group achieved better scores than the first responders, as they had the opportunity to individually initiate their thoughts and interactively accommodate and integrate other lines of thought from their dyad.

The impact of being a "one-time first responder" is even more significant when these initial contributors offer limited or inaccurate input. In contrast to the low-low dyads, high-scoring individuals in the low-high group not only engage in the elaboration phase but also progress to the negotiation and building-on phases. The negotiation phase provides them with the opportunity to accommodate differences in perspectives and ideas, fostering a more collaborative and comprehensive understanding (Hull & Saxon, 2009; Stahl, 2006, Zhu, 1996). On the other hand, the building-on phase allows for the revision and refinement of mental representation structures, leading to a deeper and more sophisticated comprehension of the subject matter. Through iterative revision and refinement, dyads construct the mental framework of the learned concepts that can be applied to other contexts (Akyol & Garrison, 2011; Schmidt & Boshuizen, 1993). These subsequent phases can even lead to score variations between dyad members, as first responders in this group tend to remain in a restricted elaboration phase while their partner advances to the negotiation and building-on phases.

In excerpt below, Participant 16 explained why people are likely to buy discounted products based on the anchor effect. Participant 20, however, elaborated on Participant 16's explanation, providing additional insights such as how people perceive the price as reasonable when they see the word "sale" on discounted products. Moreover, Participant 20 extended this notion to encompass broader constructs, such as the concept of negotiation, and applied it to various real-life instances. In this incident, Participant 16 initiated conceptual comments but missed an opportunity to interact and leverage the application-level comments provided by Participant 20. This scenario could have led to different outcomes, potentially moving this dyad from the low-high group to the high-high group if Participant 16 had revisited the thread, spent time in the build-on phase, reflected on what Participant 20 had contributed, and consolidated all levels of the learned concept.

Participant 16: "People are more likely to buy discounted goods because they are anchored in with the promise of a sale, and they see the lower price point associated with it."

Participant 20: "I agree, people see the word "sale" discount and assume that the price that they are getting is reasonable, whether it really is or not. You can see examples of anchoring in *negotiations*: people will start with a value that is much higher than the actual value of what they are selling. Then when they decrease the price, the customer thinks that they are getting a good deal, even if they are paying more than the actual value of the item. *Some examples* of these negotiations include buying a car, house, etc."

In another example, participant 116 in excerpt X provided his definition of Peak-End effect and elaborated his concept understanding as linked it to his experience. Participant 11 built on participant 116 as she extent to another line of thought describing how and why a painful experience can end on a good note and create positive experience.

She also tried to negotiate the application of Peak-End effect on a negative experience that participant 116 incorrectly introduced as she proposed an alternative explanation and clinical evidence mentioned in the video. Nevertheless, as participant 116 never revisited the discussion, he lost a chance to exercise his ideas in the other contexts, negotiate and reflect his understating with his dyad, address what is incorrect and why, and be able to revise his conceptual understanding.

The negotiation phase in collaborative learning requires group interaction to gradually align their shared understanding (Stahl, 2003). However, in an asynchronous collaborative mode, additional effort is needed to keep track of the ongoing discussion, leading participants to frequently provide minimal comments just to fulfill the task requirements (Hew et al., 2010; Hewitt, 2005). Consequently, this premature cessation of discussions may hinder the full potential of negotiation and exploration of ideas within the dyads.

Participant 116: "Peak End Effect is a phenomenon that when it comes to experience, we only remember the peak of the event and the end of the event. An example of the Peak End Effect is eating at a buffet, we only remember eating good food and walking out full. Based on Peak End Effect we should end a bad experience at the peak and end a good experience at the end. Ending with a good experience will outset the peak of a bad experience."

Participant 11: "The peak end effect is when a scenario happens and people only remember the peak and end of the event. One example of this would be when running in a long race. For instance, marathons can be painful but crossing the finish line, getting food, receiving a medal, and other benefits at the end of races make it seem like it was an amazing experience, even if the race was painful. You should end both good and bad experiences with a good experience at the end since duration does not play a role in determining a good and bad experience. People just want to end on a good note."

Participant 11: "Hi *participant 116 name*, you provided a great example for this effect! *However*, I do think ending a bad experience at the peak can be more harmful than ending it on a good note at the end. For instance, *in the video*, people in having a painful procedure

> preferred the procedure that ended less painfully than the peak compared to the procedure that ended at the peak of pain."

Low-High: Synchronous Voice-based Group (SV). In the SV, some dyads encounter challenges in comprehending the video content, requiring additional time to thoroughly digest information and construct knowledge. The collaborative nature of this mode eliminates the issue of being "one-time first responders" since dyads are concurrently engaged in discussion. However, the real-time requirement for discussing unfamiliar concepts may lead to feelings of being overwhelmed. As a result, dyads in this setting engage briefly in elaboration and attempt to describe concepts in detail. However, they soon encounter difficulties and enter the negotiation phase as they grapple with understanding unknown situations and relating them to their experiences. Although the synchronous aspect facilitates the natural progression of the negotiation phase, dyads may need more time or preparation to contemplate novel concepts fully.

In the following excerpt, Participant 17 touched upon several key points of loss aversion, but he required additional time to think, organize, and revise his discussion to effectively communicate with Participant 03. In response, Participant 03 attempted to build on Participant 17's comments by applying the concept of loss aversion to gambling and investment scenarios. However, due to her limited understanding of the concept and challenges in keeping up with Participant 17, she lost momentum and eventually gave up.

Although synchronous collaboration facilitates negotiation, a critical underlying assumption is that dyads should possess some related knowledge or basic prerequisites beforehand to effectively co-construct knowledge, as the fast-paced nature of synchronous collaboration restricts opportunities for reflection.

Participant 03: "Okay, can you actually go first this time because that whole scenario honestly may no sense."

Participant 17: "Basically, [loss aversion] referring to the natural human tendency to Really avoid loss. Even when there is a possible good game. It causes people to focus more on the negative where they lose lives, and so they choose to lose less live. Basically, people will change their perception of a better choice Depending on how the loss is described. If you're trying to sell someone something and you're focusing on the possible risks, for example. Then they're not going to be as likely to go for whatever you're trying to sell them, then, if you were trying to describe all of the different benefits for the possible benefits. *You might think not make sense, but I need more time to really described my words more clearly."*

Participant 03: "I mean like gambling it's sort of just like gamble towards that. Because you think you're going to win more. I don't know it's really hard to explain with gambling and investment because I don't have…I don't know how investment works and I don't know how to gamble so I don't know. It's sort of hard to explain."

Participant 17: "I can imagine, some people who weren't raised like I was would be more afraid of just making the right decisions if they feel that their decision might lead to loss just by action."

Participant 03: "I feel like a lot of these examples in the video are really hard to follow"

On one hand, the synchronous affordances in collaborative settings place all participants in real-time interaction, compelling them to actively contribute and minimizing the ghosting issue commonly observed in asynchronous collaborations, particularly in small groups like dyads in this study. This synchronous mode fosters a dynamic and engaging atmosphere, promoting active participation among participants.

On the other hand, novice participants may encounter challenges in maintaining focus and collaborative momentum in such a fast-paced mode. The transient nature of the talking discussion leaves no record for participants to revisit and contemplate upon (Hrastinski, 2008; Jeong & Hmelo-Silver, 2016). Unlike dyads in AT, where discussions can be revisited and reviewed at any time, dyads engaged in SV often move forward without revisiting the content once the session is completed, assuming the activity is concluded. As a result, opportunities for reflective learning and deeper understanding may be missed in the absence of revisiting and iterative co-construction.

4.8.3 High-High Group

Dyads in the high-high group predominantly originated from the SV. Compare with low-high group, the high-high group from both collaborative modes has more substantive comments. Consistent with other groups, dyads from the SV tended to generate more substantive comments compared to those from the AT within the same group. In this high-high group, the trend aligned with the general pattern, as SV dyads were more inclined to create interactive comments, while AT dyads leaned towards producing constructive comments.

Notably, while the number of differences between constructive and interactive comments in the SV was similar in the high-low and high-high groups, the discrepancies in the AT showed more variations among the low-low, low-high, and high-high groups. In the high-high group, the distinctions between constructive and interactive comments were the least pronounced among all the groups. This well-balanced proportion of constructive and interactive comments indicates dynamic exchanges of ideas and knowledge among participants, illustrating a more well-rounded interactive collaborative engagement.

High-High: Asynchronous Text-based Group (AT). Upon closer examination of dyad interaction, high-high dyads in the AT primarily engaged in elaboration and building-on processes, similar to those in low-high group. However, the interaction within high-high dyads displayed distinct characteristics, being reflective, interactive, and balanced.

For instance, the initial post by participant 25 was descriptive and reflective, elaborating on the underlying assumptions of the pricing bias concept and its applications. Subsequently, participant 115 responded meticulously, addressing participant 25's main ideas point by point. As participant 25 asserted that pricing bias involves consumers assuming a product's quality is linked to its price, participant 115 concurred and expanded on this line of thought, indicating how pricing bias can even impact products of the same quality.

When the dyads attempted to apply the learned concept to the price of wine, participant 25 further elaborated and refined her ideas, suggesting that pricing bias emerges due to the expectations associated with prices, which in turn influence subjective experiences. Participant 115 endorsed participant 25's notion and linked the ideas back to the placebo effect, which participant 25 mentioned earlier.

In contrast to the AT dyads in the low-high group, where the second responders tended to generate more substantive comments, the dyads in the high-high group interactively provided equal substantive comments. This balance and interactive engagement within high-high dyads distinguish them from dyads in the low-high group, showcasing a more collaborative and interactive approach to knowledge construction.

Participant 25: "Pricing bias is the tendency for the consumer to assume the quality of a product is tied to the price of that product. Someone might assume on brand ibuprofen (advil) works better than generic ibuprofen, for example walgreen's brand ibuprofen, because the price is higher. This bias might actually make the product more effective to an extent because of placebo.

> Expensive wine appears to taste better than cheap wine because taste, to an extent, is subjective. How much someone likes something is subjective. If someone predicts that an expensive wine must come from better grapes or have aged longer, their expectations will change their subjective experience of the taste of wine. This expectation of better or worse quality will actually affect how much they like the expensive vs cheap wine, even if the wine is the exact same."

Participant 115: "I definitely agree that pricing bias is the tendency of the consumer to buy or use a product because of their bias on the price, where the more they pay the better the product is. Another example of pricing bias that I can think of would be sunglasses, even though two sunglasses can come from the same factory, but due to branding and pricing, such as Oakley and Gucci would affects the consumer's decision to purchase the more expensive one.

> The expensive wine can appear to taste better because of a placebo effect that the higher price places on the consumer, as participant 25 said above, the higher price of the wine can come with the beliefs that the grapes that makes the wine is better, or it has been aged longer. Even if they are the exact same wine, the brain will be tricked to favor the more expensive one."

High-High: Synchronous Voice-based Group (SV). In contrast to high-high dyads in the AT, which primarily engaged in elaboration and building-on, SV dyads experienced all three phases of co-construction. The progression of knowledge co-construction from surface to deep levels demonstrates a more dynamic and interactive collaboration. Particularly in the negotiation phase, SV dyads demonstrated a wide variety of interactive discourses, including struggling, hesitating, clarifying, and proposing alternatives, as they strived to attain a mutual understanding.

Video pausing or revisiting specific portions of video content are common strategies used in annotation tools to bookmark specific information or break the video into consumable pieces. In contrast, SV dyads rarely utilize the strategy of pausing or revisiting specific portions of the video, even when they exhibit signs of struggling to comprehend the content. This reluctance to pause or revisit may be attributed to social pressure, as they interpret this action as an interruption and feel hesitant or embarrassed to delay the discussion.

However, this study found that among SV dyads, only those in a high-high group paused and revisited a video as a strategy to compensate for the limited reflective time and prepare themselves before engaging in the discussion. Their decision to pause is often triggered by internal conflicts or uncertainties in their own understanding. During these pauses, the dyads engage in self-explanation and revisit specific portions of content to solidify and ensure their conceptual understanding before proceeding with the discussion with their partners.

In the provided excerpt, participant 25 hesitated while responding to the prompt and requested her dyad to rewind the video. In contrast to the first round of passive watching, this hesitation triggered her engagement in self-explanation during the subsequent revisit, where she described contradicting scenarios caused by the mental accounting effect and proposed a hypothesis. This moment of hesitation, in turn, transitioned the dyads into a negotiation phase, characterized by high uncertainty, prompting them to collaboratively elaborate on the content, review the developed concepts, and revise their shared understanding.

The time and effort invested in the negotiation phase distinguish the SV high-high group from the high-low group. While the high-high group prolongs the discussion until they reach a shared mutual understanding, the low-high group tends to cut this phase short and prematurely end the discussion, leading to discrepancies in the understanding levels and scores within a dyad. This observation emphasizes the significance of the negotiation phase in SV interactions, as it enables dyads to actively engage in co-constructing knowledge, bridging knowledge gaps, and arriving at shared meanings.

Participant 25: "What was the example that they gave at the beginning with the movies, I didn't actually..*I just rewind and watch it again?* When we lost the second part I missed the second part of that.

yeah. *What I just explained it to you actually..* so the first one he got.. He purchased a ticket those $10 or $20 and then he lost the ticket. And then so by losing the ticket he just went home instead of buying a new one, and then in the second one, she gave him to $10 bills and he lost one of the $10 bills. But instead of going home he just purchased another ticket. So that was so in the first scenario when he lost the ticket that he bought with 10 to $10 he went home, but in the second one he lost $10 before it was a ticket you just bought another ticket so. Like he didn't feel as bad when he lost. The $10 I guess. Maybe the 10."

Participant 20: "It is like the ticket he lost would have been money he was like using for entertainment, whereas the money he lost was just his savings, maybe. Like what different accounts in his mind. So maybe since he hadn't spent it on the ticket yet he didn't consider it as like being spent on entertainment, or like something fun. Like buying another one because that's it's $10 in the like fun account in his mind.

Participant 25: "So, in the first one he already spent the money on the ticket. And so the other $10 weren't supposed to go to entertainment, maybe, so he didn't want to spend the other $10 they would have otherwise been saved for something else. So, in the second one when he loses the $10 it was going to be spent anyway, so he feels less bad about it, I think right. Okay, I think Okay.

The establishment of shared mutual understanding in the negotiation phase serves as a necessary stepping stone to reach the building-on phase. Once the dyad aligns regarding the example describing mental accounting in the video, they proceed to collaboratively elaborate on the concept. Participant 20 provides a definition of mental accounting, emphasizing its application in budgeting and expense tracking. Building on this, participant 25 further explores how people exhibit bias towards each mental account, specifically contrasting how they treat saving as opposed to cash. Participant 20 then picks up on the notions of saving and cash, delving deeper into how individuals perceive and handle redundancy within mental accounts. Subsequently, she consolidates her ideas and applies the learned concepts meticulously in the context of a movie ticket case, stating, "We wouldn't want to spend $10 on entertainment twice, but if we spend $10 on something and then spend another $10 on entertainment, that would be fine."

Participant 20: "I would say, mental accounting is in your head planning and like *keeping track of where you spend your money,* and we spend on. So maybe some examples would be like, if you know you have that you need to only spend so much money on like fun things like entertainment clothes, whatever. Then you're only get your… like think oh like I've already spent too much on this like this week, this month, whatever and not spend any more, even if you're not necessarily *like keeping an actual budget or like writing it down or anything*."

Participant 25: "It's also like keeping different separate in your mind so like. *Like money that you would save, you would treat different*… like money that you had saved instead of going to the movies, you would treat as like extra money rather than like... *like if you had that money in your savings, you could put the $10 into your savings and then it would be like you can't touch that money anymore, but if you just had the money lying around you would feel more like easy to spend* it, it would be easier to spend it on something that isn't necessary."

Participant 20: "So that's like the answer the second question. Like if you feel like you've already spent it on the ticket, you're not going to get an extra one. But if you saw like cash laying around which is like what you would consider your extra money, you're like more likely to use that.

That's why I like what you were saying about like how savings… you might be more likely or like extra money cash whatever you're more willing to spend so that's why you'd be more likely to say like spend another $10 cash. Like in the example then to buy another movie ticket, because in your mind your cash is like your extra money, whereas your ticket that you've bought is like what you sent out entertainment. *We wouldn't want to spend $10 on entertainment twice, but if we spend $10 on if we lose $10 and then spend $10 on entertainment, that would be fine."*

In addition to spending time to establish a shared understanding, SV high-high dyads utilize the negotiation phase to propose alternative views and provide constructive feedback. Peer feedback enables dyads to consider learned concepts from different perspectives, leading to new insights and supporting ongoing knowledge co-construction

(Carless, 2016). Unlike the AT, the SV allows dyads to promptly give and receive feedback that is contextually connected to the discussion of concepts. This immediacy provides dyads with the opportunity to react to the feedback received, such as revising their understanding, seeking clarification on the feedback, reflecting on the feedback, and sustaining knowledge negotiation.

In the provided excerpt, participant 16 introduced the constants in the peak-end effect (peak, end, duration) but inadequately described how these constants function and relate to each other. She mentioned that overall experience is the result of peak and end averages but failed to describe how duration moderates good and bad experiences. Participant 35 further elaborated on the implication, stating that the effect can make us feel less pain in a bad experience and happier in a good experience. Building on the implication Participant 35 provided, Participant 16 attempted to connect the peak-end effect to duration, suggesting that a longer time in a bad experience may allow the chance of a good experience to happen (e.g., "if drawn out longer, it gives the opportunity for better things to happen").

However, this notion contradicts the essence of the peak-end effect, which asserts that people only remember the peak and end of an event, and the impression of the experience is the result of the peak and end averages regardless of the duration. Participant 16's perspective seems to align with common sense, but it represents one of the common misconceptions about the Peak-End effect. She has partially grasped the essence of the Peak-End effect as explained in the video but incorrectly made inferences as she has missed the concept of duration in the video.

Participant 16's conflict arose as the experimental findings contrasted with what she observed in real life (e.g., "I guess, in theory, you wouldn't want a bad experience to last longer"). As a result, she could only remember and report what she learned from the video (e.g., "yeah, I guess a bad experience should be longer, and a good experience should be shorter") but failed to resolve the discrepancy and explain why and in what circumstances the overall experience can be altered by the sequences of sub-events, rather than the duration.

Although participant 16 engaged in active learning, her learning was limited as she merely remembered some quotes from the video instead of constructing her own understanding. Her fragmental knowledge led to misconceptions. In fact, the video explained that when the peak experience is bad, one should not end at the worst peak but should prolong the event and add good sub-events to end with a positive note. Conversely, when the peak experience is good, one should end at the best peak, so the peak and end average yield the highest result.

The SV allowed participant 35 to immediately pinpoint the potential misconception and propose different views. Her feedback also addressed participant 16's early notion of duration and reframed it to the sequences of events instead (e.g., "I think it just depends on what the peak and then the end effect, like where it is with certain samples, maybe not so much duration"). Besides conceptual reframing, she concretized her point by providing an example of the worst experience, where the peak and the bad end are the same, with no good event following (e.g., "if it were to be a very bad experience, like the peak and the end experience would be at the top, because you're not going to have a good experience right possibly the peak, so it just sucks even more, yeah, I agree"). SV enhances collaborative learning by facilitating immediate clarification and correction of misconceptions at the outset, thereby reducing cognitive load and fostering ongoing dialogue (Hrastinski, 2008; Stahl, 2003; Sweller, 2019).

Participant 16: "Peak-end effect is kind of like an average where you're taking like the peak of the event and also the end and comparing the two of them, like the highest pie, and the lowest low comparing them and, like that's also affected by duration."

Participant 35: "Like taking maybe an event and because, with the weird way that our brains work like the peak and then the end is like *what we remember maybe most about an event* so kind of base of. And then, when should you end a bad and a good experience so less pain and more happy and so like basically the surgery example is like a bad experience."

Participant 16: "If anything should be drawn out more, because then the peak doesn't look so bad if you draw it out so more i've drawn out longer I guess. Well, like or it *like gives the opportunity for better things to happen*, perhaps I just feel like experience is kind of hard to like. Make quantitative like yeah i'm having a good time versus having a bad time it's like that's hard to like put on a scale. And, like, I guess, in theory, *you wouldn't want a bad experience to last longer*, but I mean at it just with our brains it works better, but yeah I guess *a bad experience be longer and a good experience should be shorter*."

Participant 35: "um *I think it just depends on what the peak and then the end effect like where it is* with certain samples, *maybe not so much duration*, because i'm like time still.

It just kind of depends, but I just think like what the peak and effect *where those things are* so with a good experience if the peak and the end effect are like at the same like I guess plot like if it's both like the peak and the ends, then it's going to be a really good experience.

And for like a bad experience like you would probably you know you would probably want the end to be lower than the worst peak to you know.

But if you like, *if it were to be a very bad experience, like the peak and the end experience would be at the top, because you're not going to have a good experience right possible the peak so it just sucks even more yeah I agree."*

Participant 16: "yeah I agree."

In both low-high and high-high SV dyads, the negotiation phase plays a crucial role due to the heightened uncertainty. The presence of uncertainty tends to demoralize the low-high scoring group, leading them to prematurely end discussions and fail to establish a shared understanding. This premature ending and failure to establish shared understanding in high uncertainty scenarios may leave one participant behind, resulting in a low-high dyad, as observed in the case of participants 03 and 17. In contrast, the high-high scoring group perceives uncertainty as a valuable learning resource, prompting them to collaboratively engage in inquiry, exploration, and explanation, aligning with the practical inquiry model of cognitive presence (Akyol & Garrison, 2011; Caskurlu et al., 2021; Chen & Techawitthayachinda, 2021).

CHAPTER 5

DISCUSSION AND IMPLICATIONS

Implication for Future Research

This study extends the applicability of social constructivism to video-based learning through collaborative activities and communication modes. Collaborative video viewing is a social constructivist learning activity that promotes individuals in constructing their understanding and exchanging interpretations, perspectives, and insights with their peers. This aligns with the core principle of co-constructing knowledge through dialogue and interaction, a fundamental aspect of social constructivism (Cress & Kimmerle, 2008; Kalina & Powell, 2009). Situating social constructivism in comparative collaborative environments, this study emphasizes the impact of technological affordances in the collaborative learning process, transforming video-based learning from a passive experience into an active endeavor. In line with previous literature that highlights the role of the social dimension in enhancing learning outcomes (e.g., Chi, 2008; Stahl, 2006; Vygotsky; 1978), this study provides the contextual boundaries of the group interaction process in online collaborative video viewing, offering insights for future research and implications for instructional design in active video-based learning.

5.1 The Effect of Collaborative Modes on ICAP Learning Engagement

When comparing the synchronous voice-based (SV) and asynchronous text-based collaborative modes (AT), the former group performs better in application scores. Higher application scores suggest that synchronous voice-based dyads excel in achieving transferability, deepening their understanding of concepts, and applying their knowledge effectively in practical and relevant ways across different situations.

When examining learning engagement during group interactions, the study's findings are consistent with the ICAP model's assumption that the quality of learning is influenced by the number of overt activities and the associated cognitive engagement levels. These levels are listed as Passive (P) < Active (A) < Constructive (C) ≤ Interactive (I) (Chi & Wylie, 2014). This study found that the collaborative mode had an impact on learning engagement in terms of both quantity and order. Specifically, SV dyads tended to produce more substantive comments than AT dyads. In terms of the ICAP order, SV groups leaned toward interactive engagement, while AT dyads demonstrated constructive engagement. Furthermore, in line with ICAP predictions, within the same collaborative mode, high-high groups generated more substantive comments compared to low-high, average, and low-low groups, respectively. This confirms the ICAP model's predictions in online interactions and extends its applicability beyond the originally developed framework from face-to-face environments.

However, it's important to note that this prediction may not hold across different collaborative modes. Despite producing a similar number of substantive comments, the high-high AT dyads outperformed their SV counterparts in the average and low-high groups. One explanation is that the higher generative contribution related to application prompts in high-high AT dyads makes them outperform their counterparts. The highest level of generative comments in application prompts is also observed in high-high SV dyads compared to other dyad-score groups in the same and different collaborative modes. This underscores the significance of coding for more than just substantive or content-related comments but focuses on the generative process involving contributions beyond mere repetition of the video content and what their partner mentioned. Another plausible explanation is that the nature of the textual modality allows for more conscious, concise, and accurate responses (Hew & Cheung, 2013; Sherblom, 2010; Stockwell, 2010). The concise communication from high-high dyads in the AT may encapsulate a deeper level of knowledge. It could involve several abstract, higher-order concepts, simplified causal mechanisms, or succinct summaries of a larger number of lower-level concepts (Schmidt & Mamede, 2020). This observation aligns with a key point highlighted in Kent et al.'s (2016) research, where they established a correlation between the quantity of contributions and their quality. However, they caution that the quantitative data alone, such as the number of interactivities, does not capture the full context or connectedness and recommend further analysis of knowledge depth for future research. Their argument also resonates with the current study, where high-high SV dyads created the fewest turns among other SV dyad-score groups, but each turn consisted of the most generative comments, indicating a continuous and descriptive flow of knowledge development.

When comparing the proportion of constructive and interactive comments, SV interactions showed a more balanced ratio compared to AT. This balance suggests a more collaborative effort. However, even though generative comments in AT were fairly evenly distributed, with the exception of the low-low group, interactive comments set different groups apart, resulting in high-high, average, and low-high groups. This imbalance indicates more divergent contributions between dyads, leaning more towards individual rather than collaborative activities. In fact, the wider gap suggests a tendency to achieve lower dyad scores, such as low-low or low-high, indicating lower interactive engagement and a potential passive stance from one or both partners.

These findings indicate that AT dyads are primarily engaged in constructive comments while limiting their interactive comments. This suggests two strategies for CSCL scripts: a) employing epistemic scripts to empower knowledge construction (i.e., increasing C) and b) utilizing social scripts to compensate for the deficiency (i.e., increasing I). CSCL scripts encompass a variety of strategies, including specifying and sequencing learning activities, as well as focusing on specific discourse to develop internal scripts (Schnaubert & Vogel, 2022). Weinberger et al. (2005) have addressed group interaction issues in computer-supported collaborative learning (CSCL), both in SV (video conferencing) and AT (discussion boards), and thus propose using epistemic and social scripts to enhance collaborative learning. In their study, epistemic scripts are designed to guide learners' attention by providing prompts developed based on expert task strategies, while social scripts assign student roles as constructive critics. However, their findings reveal that epistemic scripts did not effectively support learners in developing their own conceptual understanding. In fact, epistemic scripts may hinder collaborative efforts by reducing the need for integrating diverse ideas and interactive engagement. On the other hand, social scripts that specify effective interaction patterns reinforce the collaborative learning mechanism and help learners evaluate and refine their conceptual models by integrating various perspectives. However, Zhang et al. (2020) argued that CSCL scripts and designs should facilitate "climbing the social and epistemic ladders together." They found that the "super note" feature in Knowledge Forum, enabling the horizontal integration of ideas, increases social interaction. This, in turn, allows for vertical moves in problem exploration.

5.2 The Affordances of Collaborative Modes on Group Interaction

Based on the quantitative findings, this study demonstrates that collaborative modes influence learning engagement and, consequently, may impact learning outcomes. In this section, we will further discuss how the affordances of collaborative modes, stemming from social engagement, cognitive processing, and information processing, may influence group interactions.

5.2.1 Social Engagement

One potential explanation for the higher number of generative comments and improved application scores observed in SV is that the enhanced social engagement facilitated by SV interactions encourages more generative processes. Having a partner in rich media environments plays a crucial role in promoting social engagement and motivation (Sherblom, 2010). Furthermore, the social demand inherent in SV interactions can motivate passive participants to invest more effort in the collaborative process (Chen et al., 2007; Asterhan & Schwarz, 2010). The presence of a peer may alternate the effect of self-explanation and explanation to others (Roscoe & Chi, 2008). In a study by Pi et al. (2022) examining the presence of a co-learner while watching a video, they found that even without direct peer interaction, the mere presence of a co-learner group resulted in higher-quality explanations, reduced learners' mental effort, and led to more behaviors related to self-regulation and monitoring, as compared to scenarios with no co-learner.

However, the enhanced generative processes are due to increased group interaction facilitated by social presence, not just the presence of a peer. This distinction becomes evident when considering the various score groups observed in SV. Supporting this notion, a study by Liu et al. (2023) found that students who watched videos alone achieved similar results to those who co-watched without interaction, while co-watching with interaction led to better learning performance and metacognition. Additionally, Giesbers et al. (2013) discovered that in an online course, the use of videoconference tools correlated with motivation and final exam scores. However, they emphasized that group interaction appeared to be a stronger predictor of final exam scores than tool usage. Wang and Wang (2019) also found that collaborative modes directly affect social presence which they found lay ground for cognitive presence, but not learning outcomes. These findings suggest that collaborative modes can potentially influence group interaction during collaborative video viewing but might not directly impact learning outcomes.

In contrast, dyads in an asynchronous text-based collaborative mode are less exposed to social involvement, which results in group interaction having a lesser influence on learning outcomes and placing a greater emphasis on individual effort. This assertion aligns with observation that their learning process tends to involve more individually constructive engagement than collaboratively interactive engagement. The passive encounter nature of dyads in asynchronous mode sometimes causes them to assume a passive role, consistent with the long-standing and intractable issues in the online learning literature (Andresen, 2009; Hewitt; Warren, 2008). Some passive behaviors such as adopting a 'wait and see' strategy or, in severe cases, 'thread dead'—ghosting their partner's contribution—are also observed in this study.

5.2.2 Cognitive and Information Processing

Modality affordances in SV facilitate dynamic discussions, making it (a) easy to generate output (verbal affordance) and (b) engageing in spontaneous interaction (synchronous affordance) (Hrastinski, 2008; Watt, 2016). The reduced filtering and immediate interaction in voice-based interactions promote dynamic discussions, encouraging dyads to contribute more while negotiating their understanding to co-construct shared meaning (Guo et al., 2021; Park & Bonk, 2007; Molnar & Kearney, 2017). Park and Bonk (2007) discovered that while time-delayed discussions in text-based asynchronous collaborative tools can create conditions that encourage thoughtful conversations, they can also lead some students to become overly cautious when sharing their ideas, resulting in procrastination. In contrast, in voice-based synchronous collaboration, students are more inclined to take initiative and are more willing to experiment with their ideas. In Guo et al. (2021) study, students similarly expressed that they found speaking to be more natural, whereas they had concerns about the language and formality of the discussion discourse in discussion boards, which inhibited their ability to express their thoughts.

However, due to the ease of output production, similar to speaking in a face-to-face environment, the speaking output in synchronous voice-based communication tends to be short, fragmented, unorganized, and incomplete (Schneider et al., 2002; Stockwell, 2010). To compensate for the lack of reflective time in a fast-paced synchronous collaborative mode, dyads briefly internalize and create mediocre output while shifting their focus to rely on external resources from their partner to revise their internalization, as evidenced by their more detailed, complete contributions as discussions progress. The increased active momentum within the dyad provides them with additional input from their interactions to further internalize and, in turn, revise their externalization, connecting internal and external loops and enhancing the richness of the dual loop of co-construction. This finding underscores the crucial interplay between individual cognition (the private world) and collective understanding (the shared world). It highlights the vital role of group interaction, aligning with established collaborative models such as the conversational framework (Laurillard, 2007), knowledge negotiation (Stahl, 2006), and the practice of inquiry model (Garrison et al, 2010).

Expanding upon the concept of the dual loop of co-construction, in high-uncertainty situations during the negotiation phase, the spontaneous interaction in synchronous voice-based dyads derive advantages from the collective memory effect. This effect enables them to unload their individual memory demands and rely on mutual processing. Kirschner et al. (2018) view collective memory as the outcome of larger and more efficient cognitive processors. In collaborative settings, the processing load is distributed among group members, leading to a reduction in cognitive load.

However, this study has revealed that the swift pace of synchronous collaborative video viewing can impose a significant cognitive load on individuals with lower proficiency. They are tasked with immediately processing all cognitive resources, whether from the video or their dyad partner (i.e., high components in cognitive load theory, Sweller, 2020). In this context, novice participants with less structured mental representations may find that relying on external resources becomes counterproductive.

The cognitive load associated with processing external information and the transition cost of communication often outweigh the benefits of collective memory, leaving them with diminished capacity to internalize their understanding (e.g., 'Um... I am not quite sure. Let me think for a second.'). Consequently, shared understanding is not always achieved, creating a gap and resulting in dyads with varying levels of comprehension. This highlights the conditional interplay between learner proficiency and collaborative modes in CVV, which offers fertile ground for further exploration in future research.

In asynchronous mode, the cognitive load associated with managing multiple components is reduced due to the time delay. However, this trade-off of the time delay is the lack of spontaneous interaction, which may hinder the flow of co-construction. SV dyads demonstrate a gradual alternation of externalization and internalization, eventually developing a shared understanding through multiple turns of dialogue, whereas AT dyads tend to show a preference for greater internalization, engaging in self-explanation (Chi et al., 1989). If interaction occurs among AT dyads, it typically involves a single dyadic turn, consisting of initial input and response. Chou (2002) also reported similar findings, noting that asynchronous students predominantly engage in one-way communication with uneven interaction, where most students' comments do not necessitate further clarification or stimulate further discussion. In contrast, students in the synchronous communication mode engage in more spontaneous back-and-forth communication, with contributions more equally distributed.

The relatively low level of social involvement in AT interactions has the potential to negatively impact cognitive engagement (Garrison et al., 2010). This is of particular concern when cognitive engagement is already limited, and the opportunity for social engagement to address cognitive issues is omitted. This aspect becomes evident in the current study, where low-low and low-high dyads exhibit minimal social engagement. These findings underscore the intricate interplay between social interactions and cognitive processes in collaborative learning.

5.3 Co-Construction Process in Collaborative Modes

The findings revealed that dyads in both collaborative modes employed a variety of similar interactive discourses during the co-construction process. The shared discourses include 'elaborate,' 'explain,' 'hesitate,' 'alternative,' 'build on,' and 'challenge.' However, some discourses exclusive to SV dyads are 'struggle' and 'clarify.' These exclusive discourses suggest that SV dyads may experience more uncertainty and attempt to collaboratively regulate and negotiate shared meanings to reduce uncertainty and revise their conceptual understanding through discourses.

These discourses demonstrate both shared and unique interactive patterns, including the probability of a specific discourse being elicited by an initial discourse. In this study, it was found that over 80% of the time, both SV and AT dyads engage in sequences such as 'elaborate' → 'elaborate,' 'elaborate' → 'build on,' and 'explain' → 'elaborate.' Additionally, unique patterns were observed in SV dyads, including sequences where the initial comment is associated with 'struggle,' 'build on,' and 'clarify,' such as 'struggle' → 'elaborate.' Regardless of the nature of the initial comments, the two-gram sequential analysis showed that both SV and AT dyads tend to respond with 'elaborate' and 'build on' to further co-construct knowledge. The interactive discourses observed within group interactions further be be classified into three phases: co-explanation, negotiation, and application, based on their functional similarities.

Due to the fluidity of modality and the high level of dynamic interaction, dyads in SV, generally both in low-high and high-high groups, tend to engage in negotiation and a building-on phase more than AT dyads. This is consistent with previous studies that demonstrate the association between knowledge negotiation and higher-order thinking. Garrison et al. (2001) propose a practical inquiry model that incorporates cognitive presence into four phases: (a) triggering event, (b) exploration, (c) integration, and (d) resolution. These phases can be mapped onto co-construction phase identified in the current study: co-explanation phase (trigger event-exploration), negotiation phase (exploration-integration), and application phase (integration-resolution).

Both SV and AT dyads across different score groups shared involvement in co-explanation, but they differed in the extent to which SV and the higher-scoring groups tend to favor it. This is consistent with the practical inquiry model and previous studies, which indicate that students generally perform well when exploring, generating, and sharing ideas but often face challenges when it comes to reconciling ideas and concluding solutions (Garrison, 2001; Sun et al., 2022, Zhang et al., 2022, Zhu et al., 2019).

However, the negotiation phase, which occurs in around 20% of all discussions in SV and 5% in AT, may distinguish low-high dyads from high-high dyads. The ability to reconcile differences and construct shared meaning is one of the key components that differentiate low-high from high-high dyads. Our findings demonstrate that, in the defined negotiation phase, low-high SV dyads tend to remain in the exploration process and fail to reach a shared understanding. In contrast, high-high dyads progress to the integration process, where they establish the necessary shared conceptual foundation to move forward to the application phase. As the issues that arise in the negotiation phase are usually related to structure rather than details, this phase is critical for developing a shared conceptual understanding within the group and for being able to apply ideas to other contexts in the application phase. This aligns with Zhu et al.'s (2019) findings, which emphasize a critical point in the negotiation process. In successful tasks, students ensure everyone shares an understanding of variable relationships before moving forward, while in unsuccessful tasks, the process is often regulated without achieving shared understanding.

Contrary to the negotiation phase in SV, this study did not find a negotiation phase in high-high AT dyads. However, we observed some one-way negotiation discourses in low-high AT dyads initiated by the higher-scoring partner. The absence of a negotiation phase in high-high AT dyads can be attributed to the already completed contributions from both partners, which are typically reflective, extensive, and well-formed. They often bypass the negotiation phase, moving directly from co-explanation to the application phase. In contrast, Zhang et al. (2022) compared low and high-performance students in collaborative problem solving on an instructional design task in asynchronous text-based collaboration. They found that proportionally, low-performance students have a higher frequency of sharing ideas, whereas high-performance students engage in a higher degree of negotiation discourse involving the discussion of different ideas. This discrepancy of negotiation phase in high-score asynchronous text-based group between this study and Zhang et al. (2022) may stem from the fact that they used a variety of asynchronous text-based collaboration tools, each with its unique design and affordances, while the current study utilized a social annotation tool. This highlights the importance of considering collaborative environments beyond collaborative modes and taking into account the ecological design of technology and pedagogy (Smith et al., 2003).

On the other hand, in low-high AT dyads, one-way negotiation often involves the higher-score partner proposing alternative explanations and providing constructive feedback to the lower-score partner, who may have composed an incorrect or incomplete contribution. The lower-score partner in AT in this group typically adopts a 'one-time first responder' strategy, meeting the minimum requirement of answering prompts before leaving the screen. As a result, the benefits derived from the negotiation phase vary for SV and AT low-high dyads.

In SV, where synchronous affordance demands both partners to be present at the same time, both partners mutually benefit from self- and co-regulation, developing metacognition to address their own and the group's knowledge status and working to align their understanding (citing Zimmerman). On one hand, the lower-score partner benefits from peer constructive feedback and has the potential to move themselves into the high-high score group, as we also observed similarities in the initial negotiation phase of high-high SV dyads. On the other hand, the higher-score partner gains the 'explain-to-other' effect (Roscoe & Chi, 2008) by adopting a peer-teaching strategy to elaborate, explain, propose alternative views, as constructive feedback, rather than directly evaluating what is incorrect. However, in low-high AT dyads, as the lower-score partner disappears, the higher-score partner, who has invested time and effort in addressing conceptual gaps, enters the negotiation phase alone but still benefits from self-regulation and the 'explain-to-other' effect.

Limitations & Future Directions

The results should be interpreted in light of several limitations. First, while this study had a sample size based on G*Power calculations, it remains relatively small, which could potentially undermine the study's internal and external validity. Additionally, the study was conducted in a laboratory setting, which may not fully reflect how real-world students interact in collaborative modes. The fact that participation in the lab study was voluntary and carried no score punishment might have influenced motivation and the level of investment in group interaction. There may have also been some social bias at play, as SV dyads were aware that their faces and voices were being recorded, which could have motivated them to perform better. In contrast, AT dyads were not required to reveal many social traits during the discussion, such as their identity or face, which could have had the opposite effect compared to SV dyads.

The study duration is another important factor to consider. In this study, each dyad had limited exposure to both AT and SV sessions. While SV imitates a face-to-face dialogue style, AT requires more technological competency and may not feel as natural, requiring time to become familiar. For example, Akyol & Garrison (2011) found that AT dyads had low contribution and low metacognition in the first week but showed significant improvement in both quantity and quality by the ninth week.

Additional qualitative analyses, such as multiple case studies focusing on specific participants and how of knowledge trajectories developed through group interaction, offer a valuable opportunity to explore the intricacies of how individuals approach their learning within various collaborative modes. Notably, no low-low dyads were found within SV in this study. This suggests that introducing dynamic discussions within this mode may lead low-low dyads to transition to at least average dyads. This hypothesis can form the basis for qualitative analyses of knowledge trajectories, showing the progression from passive to active engagement. Single coding, where a sole researcher is responsible for data analysis, is another limitation of this study. Involving multiple coders could enhance analysis reliability and validity by offering diverse perspectives and reducing the impact of individual interpretation.

In this study, we employed a specific learning activity involving story problems within application questions. These questions were strategically designed to place concepts within real-life contexts, requiring participants to grasp underlying principles and apply conceptual knowledge demonstrated in video examples to solve problems that shared the same structure but had dissimilar features. These questions were intentionally designed as well-structured problems, each with a specific set of correct explanations. However, it's important to recognize that CVV environments offer a vast landscape for exploring diverse types of problems and problem spaces. Dyads engaged in collaborative learning may exhibit varying dynamics when confronted with ill-structured problems, spanning categories such as design problems, troubleshooting problems, and ethical dilemmas (Jonassen, 2000). For example, design problems often benefit from a reflective learning environment (Hong & Choi, 2015; Zhang et al., 2022), particularly in asynchronous collaborative modes where students often undergo multiple rounds of research and discussion before collectively arriving at a design decision.

A closely intertwined factor with problem structure in this study is problem difficulty. Significant differences were observed in application scores between SV and AT, particularly in Overall and Perception Bias videos. However, such discrepancies were not evident in the Pricing Bias video. Upon closer scrutiny, it became apparent that mean scores for pricing bias were statistically lower than those observed in the other tests. This observation suggests a potential interaction effect between task difficulty and collaborative mode on application scores. This hypothesis finds resonance in a study conducted by Jacob et al. (2020), where they explored a similar concept. Their research investigated the interaction between text complexity and communication modality (e.g., explaining orally or in written form) on comprehension levels of explanation. The findings from their study highlighted the influence of these interacting factors on comprehension.

Building on these limitations, future research endeavors can delve deeper into examining (a) classroom online collaboration, (b) the study duration and treatment exposure, (c) different analysis approach (d) problem structure, and (e) problem difficulty within the context of Collaborative Virtual Video (CVV). Such investigations hold promise for enhancing our understanding of the nuanced dynamics at play during video-based collaborative learning and can contribute significantly to the field of virtual education.

Conclusion

Video-based learning is a widely used method in online education. Lecture videos offer the advantages of high-quality instruction on demand, improved learning experiences, and scalability while remaining cost-effective (Costley et al., 2021; Hansch et al., 2015; Wong et al., 2022). However, one of the significant criticisms of online video-based learning is its passive nature, which primarily involves knowledge absorption (Chi et al., 2017; Zhang et al., 2020).

Collaborative Video Viewing (CVV) has been introduced to enhance passive video-based learning. Traditionally, CVV has been implemented in asynchronous text-based collaborative modes (AT). While the potential of synchronous voice-based collaborative mode (SV) is on the rise in current online learning, no study has directly compared the influence of collaborative modes and tools on learning outcomes, group interaction, and the co-construction process.

The key findings suggest that, in general, SV dyads performed better on application scores. When considering dyad-score groups, we found that most SV dyads fell into the average and high-high score groups. However, both SV and AT dyads were equally represented in mixed-performance groups, such as the low-high group. The low-low group consisted only of AT dyads, highlighting the persistent issue of passive interaction in asynchronous collaboration.

ICAP hypotheses were tested to verify whether the differences in mean application scores between SV and AT could be attributed to the varying levels of learning engagement influenced by collaborative modes. The findings demonstrated that SV dyads predominantly used interactive comments, whereas AT dyads tended to rely on constructive comments. This suggests that collaborative modes exhibit a tendency toward fostering collaboration in SV and self-explanation in AT. The learning engagement within the same collaborative mode aligns with ICAP predictions. Specifically, higher-score groups exhibit a higher level of learning engagement. For instance, high-high SV dyads had the highest number of generative comments (i.e., the sum of constructive and interactive comments), surpassing the low-high and low-low groups, in order. Similarly, high-high AT dyads demonstrated the same trend. However, it's important to approach the comparison of learning engagement across different modes with caution, as the differences in the structure of the outputs may affect the validity of the comparison.

Upon closer examination of their interactions, we observed that SV dyads exhibited a wider variety of interactive discourses. However, both SV and AT dyads predominantly utilized elaboration and explanation to convey their ideas, along with building on to apply learned concepts to different contexts. Notably, the struggleless and the need for clarification discourses were exclusive to SV interactions. Although these discourses indicating uncertainty occurred less frequently than the major discourses such as elaboration, the patterns in the discourse and their sequential probability revealed that they led to more extensive elaboration and subsequent building on of ideas.

Knowledge co-construction phases were identified based on the discourse and discourse patterns found, which include co-explanation, negotiation, and application. We observed that all dyad-score groups engaged in the co-explanation phase, where they described concepts, shared knowledge, and explored ideas (Garrison, 2001; Sun et al., 2022; Zhang et al., 2022). However, the extent and quality of engagement in this phase varied among the different groups.

We found that the effort dyads put into the negotiation phase could potentially facilitate the transition of the lower-score partner from the low-high group to the high-high group. However, the underlying problem that hinders this transition differs between SV and AT. In SV, lower-score partners with deficits in related concepts often struggle to process video content synchronously while engaging in a discussion. This challenge transforms useful cognitive resources into cognitive load and disrupts the double-loop process of internalization and externalization in the collaborative model (Stahl, 2006). When the lower-score partner gives up or the dyads fail to align their shared understanding, the tendency to diverge in scores increases. However, the fragmented mental representation of the lower-score partner can either (a) prevent, (b) be amended, or (c) be repaired during the pre-discussion, during discussion, or even after the discussion respectively, as discussed in the implications for instructional design. In low-high AT dyads, the issue tends to be more related to social aspects as the lower-score partner takes a passive stance and dismisses collaborative efforts, including opportunities to socially amend the fragmented, incomplete mental representation.

This study has made efforts to anticipate and address factors that could threaten internal and external validity. Nevertheless, the results should be interpreted in the context of several limitations, including the number of participants, the motivation of participants in the lab study, and the study duration and treatment exposure. Future research could explore different collaborative SV and AT tools, as these tools may significantly impact the technology-pedagogy ecosystem (Smith et al., 2003). Additionally, investigating problem spaces, including different problem types and difficulties, in the online CVV arena could enhance our understanding of learning engagement in group interaction and the co-construction process.

REFERENCES

Agarwala, M., Hsiao, I., Chae, H. S., & Natriello, G. (2012). Vialogues: Videos and Dialogues Based Social Learning Environment. *2012 IEEE 12th International Conference on Advanced Learning Technologies*, 629–633.

Aiello, G., & Thurlow, C. (2006). Symbolic Capitals: Visual Discourse and Intercultural Exchange in the European Capital of Culture Scheme. *Language and Intercultural Communication*, *6*(2), 148–162.

Akyol, Z., & Garrison, D. R. (2011). Assessing metacognition in an online community of inquiry. *The Internet and Higher Education*, *14*(3), 183–190.

An, Y.-J., & Frick, T. (2006). Student Perceptions of Asynchronous Computer-Mediated Communication in Face-to-Face Courses. *Journal of Computer-Mediated Communication: JCMC*, *11*(2), 485–499.

Andresen, M. A. (2009). Asynchronous discussion forums: Success factors, outcomes, assessments, and limitations. *Journal of Educational Technology & Society*, *12*(1), 49–257.

Asterhan, C. S. C., & Schwarz, B. B. (2010). Online moderation of synchronous e-argumentation. *International Journal of Computer-Supported Collaborative Learning*, *5*(3), 259–282.

Baker, A. (2016). Active Learning with Interactive Videos: Creating Student-Guided Learning Materials. *Journal of Library and Information Services in Distance Learning*, *10*(3-4), 79–87.

Boronat, F., Montagud, M., Salvador, P., & Pastor, J. (2021). Wersync: A web platform for synchronized social viewing enabling interaction and collaboration. *Journal of Network and Computer Applications*, *175*, 102939.

Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-03-0125

Branon, R. F., & Essex, C. (2001). Synchronous and asynchronous communication tools in distance education. *TechTrends*, *45*(1), 36.

Cadiz, J. J., Balachandran, A., Sanocki, E., Gupta, A., Grudin, J., & Jancke, G. (2000). Distance learning through distributed collaborative video viewing. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, 135–144.

Carless, D. (2016). Feedback as Dialogue. In *Encyclopedia of Educational Philosophy and Theory* (pp. 1–6). Springer Singapore.

Caskurlu, S., Richardson, J. C., Maeda, Y., & Kozan, K. (2021). The qualitative evidence behind the factors impacting online learning experiences as informed by the community of inquiry framework: A thematic synthesis. *Computers & Education*, *165*, 104111

Caspi, A., Gorsky, P., & Privman, M. (2005). Viewing comprehension: Students? learning preferences and strategies when studying from video. *Instructional Science*, *33*(1), 31–47.

Chan, J. W. W., & Pow, J. W. C. (2020). The role of social annotation in facilitating collaborative inquiry-based learning. *Computers & Education*, *147*, 103787.

Chen, H.-T. M., & Thomas, M. (2020). Effects of lecture video styles on engagement and learning. *Educational Technology Research and Development: ETR & D*, *68*(5), 2147–2164.

Chen, Y., Liu, C., & Wong, R. (2007). The adoption of synchronous and asynchronous media in the teaching of a second language. *Issues in Information Systems*, *3*(1), 217–223.

Chen, Y. C., & Techawitthayachinda, R. (2021). Developing deep learning in science classrooms: Tactics to manage epistemic uncertainty during whole-class discussion. *Journal of Research in Science Teaching*, *58*(8), 1083-1116. https://doi.org/10.1002/tea.21693

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105.

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.

Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns in peer collaboration that promote learning. *Socializing Intelligence through Academic Talk and Dialogue*, 263–274.

Chi, M. T. H., & Ohlsson, S. 2005. Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*: 371-399. New York: Cambridge University Press.

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

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243.

Cho, Y. H., & Lee, S. E. (2013). The role of co-explanation and self-explanation in learning from design examples of PowerPoint presentation slides. *Computers & Education*, *69*, 400–407.

Chou, C. C. (2002, January). A comparative content analysis of student interaction in synchronous and asynchronous learning networks. In *Proceedings of the 35th annual Hawaii international conference on system sciences* (pp. 1795-1803). IEEE. https://doi.org/10.1109/hicss.2002.994093

Choi, H., & Hur, J. (2023). Passive participation in collaborative online learning activities: A scoping review of research in formal school learning settings. *Online Learning*, *27*(1). https://doi.org/10.24059/olj.v27i1.3414

Conole, G., & Dyke, M. (2004). What are the affordances of information and communication technologies? *ALT-J*, *12*(2), 113–124.

Costley, J., Fanguy, M., Lange, C., & Baldwin, M. (2021). The effects of video lecture viewing strategies on cognitive load. *Journal of Computing in Higher Education*, *33*(1), 19–38.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 105–122.

Crook, C., & Schofield, L. (2017). The video lecture. *The Internet and Higher Education*, *34*, 56–64.

de Koning, B. B., Hoogerheide, V., & Boucheix, J.-M. (2018). Developments and Trends in Learning with Instructional Video. *Computers in Human Behavior*, *89*, 395–398.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, *46*(1), 6–28.

Deiglmayr, A., & Spada, H. (2010). Collaborative problem-solving with distributed information: The role of inferences from interdependent information. *Group Processes & Intergroup Relations: GPIR*, *13*(3), 361–378.

Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers & Education*, *78*, 312–320.

Deng, R., & Gao, Y. (2023). Using Learner Reviews to Inform Instructional Video Design in MOOCs. *Behavioral Sciences*, *13*(4), 330. https://doi.org/10.3390/bs13040330

Dhawan, S. (2020). Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22.

Dillenbourg, P. (1999). *What do you mean by collaborative learning? Collaborative learning: Cognitive and computational approaches.* Amsterdam, NL: Pergamon Elsevier Science.

Dirkx, K. J. H., Skuballa, I., Manastirean-Zijlstra, C. S., & Jarodzka, H. (2021). Designing computer-based tests: design guidelines from multimedia learning studied with eye tracking. *Instructional Science*, *49*(5), 589–605.

Erlandson, B. E., Nelson, B. C., & Savenye, W. C. (2010). Collaboration modality, cognitive load, and science inquiry learning in virtual inquiry environments. *Educational Technology Research and Development: ETR & D*, *58*(6), 693–710.

Evi-Colombo, A., Cattaneo, A., & Bétrancourt, M. (2020). Technical and pedagogical affordances of video annotation: A literature review. *Journal of Educational Multimedia and Hypermedia*, *29*(3), 193–226.

Evi-Colombo, A., Cattaneo, A., & Bétrancourt, M. (2023). Procedural knowledge acquisition in a second-year nursing course. Effectiveness of a digital video-based collaborative learning-by-design activity using hypervideo. *Journal of Computer Assisted Learning*, *39*(2), 501–516.

Fang, J., Wang, Y., Yang, C.-L., Liu, C., & Wang, H.-C. (2022). Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proc. ACM Hum.-Comput. Interact.*, *6*(GROUP), 1–21.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741.

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology*, *112*(5), 895.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415.

Fu, E. L. F., van Aalst, J., & Chan, C. K. K. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 441–478.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *The American Journal of Distance Education*, *15*(1), 7–23.

Garrison, D. R., Anderson, T., & Archer, W. (2003). A theory of critical inquiry in online distance education. *Handbook of Distance Education*, *1*(4), 113–127.

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, *13*(1), 5–9.

Geerts, D., Vaishnavi, I., Mekuria, R., van Deventer, O., & Cesar, P. (2011). Are we in sync? synchronization requirements for watching online video together. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 311–314.

Gibbons, J. F., Kincheloe, W. R., & Down, K. S. (1977). Tutored videotape instruction: a new use of electronics media in education. *Science*, *195*(4283), 1139–1146.

Giesbers, B., Rienties, B., Tempelaar, D., & Gijselaers, W. (2014). A dynamic analysis of the interplay between asynchronous and synchronous communication in online learning: The impact of motivation. *Journal of Computer Assisted Learning*, *30*(1), 30–50.

Girasoli, A. J., & Hannafin, R. D. (2008). Using asynchronous AV communication tools to increase academic self-efficacy. *Computers & Education*, *51*(4), 1676–1682.

Grabowski, J. (2010). Speaking, writing, and memory span in children: output modality affects cognitive performance. *International Journal of Psychology: Journal International de Psychologie*, *45*(1), 28–39.

Greenhow, C., Graham, C. R., & Koehler, M. J. (2022). Foundations of online learning: Challenges and opportunities. *Educational Psychologist*, *57*(3), 131–147.

Griffiths, M., & Graham, C. (2010). Using asynchronous video to achieve instructor immediacy and closeness in online classes: Experiences from three cases. *International Journal on E-Learning*, *9*(3), 325–340.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, *1*(2), 147–166.

Guo, C., Shea, P., & Chen, X. (2022). Investigation on graduate students' social presence and social knowledge construction in two online discussion settings. *Education and Information Technologies*, *27*(2), 2751–2769.

Habes, M., Ali, S., Salloum, S. A., Elareshi, M., & Ziani, A.-K. (2020). Digital Media and Students' AP Improvement: An Empirical Investigation of Social TV. In 2020 *International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)* (pp. 1–6). Sakheer, Bahrain: IEEE. doi:10.1109/3ICT51146.2020.9311941.

Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., & Schmidt, J. P. (2015). *Video and Online Learning: Critical Reflections and Findings from the Field*. https://doi.org/10.2139/ssrn.2577882

Henri, F. (1992). Computer Conferencing and Content Analysis. *Collaborative Learning Through Computer Conferencing*, 117–136.

Henriksen, D., Creely, E., & Henderson, M. (2020). Folk pedagogies for teacher transitions: Approaches to synchronous online learning in the wake of COVID-19. *Journal of Technology and Teacher Education*, *28*(2), 201–209.

Hew, K. F., & Cheung, W. S. (2013). Audio-based versus text-based asynchronous online discussion: two case studies. *Instructional Science*, *41*(2), 365–380.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58.

Hew, K. F., Cheung, W. S., & Ng, C. S. L. (2010). Student contribution in asynchronous online discussion: a review of the research and empirical exploration. *Instructional Science*, *38*(6), 571–606.

Hew, K. F., Lan, M., Tang, Y., Jia, C., & Lo, C. K. (2019). Where is the "theory" within the field of educational technology research? *British Journal of Educational Technology: Journal of the Council for Educational Technology*, *50*(3), 956–971.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *The journal of the learning sciences*, *14*(4), 567-589.

Hong, Y. C., & Choi, I. (2019). Relationship between student designers' reflective thinking and their design performance in bioengineering project: Exploring reflection patterns between high and low performers. *Educational technology research and development*, *67*, 337-360.

Hrastinski, S. (2008). Asynchronous and synchronous e-learning. *Educause Quarterly*, *31*(4), 51–55.

Hsiao, I.-H., Chae, H. S., Malhotra, M., Baker, R., & Natriello, G. (2014). Exploring Engaging Dialogues in Video Discussions. *Educational Data Mining 2014*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.660.2165&rep=rep1&type=pdf

Huei-Tse Hou, Kuo-En Chang, & Yao-Ting Sung. (2008). Analysis of Problem-Solving-Based Online Asynchronous Discussion Pattern. *Journal of Educational Technology & Society*, *11*(1), 17–28.

Hull, D. M., & Saxon, T. F. (2009). Negotiation of meaning and co-construction of knowledge: An experimental analysis of asynchronous online instruction. *Computers & Education*, *52*(3), 624–639.

Jacob, L., Lachner, A., & Scheiter, K. (2020). Learning by explaining orally or in written form? Text complexity matters. *Learning and Instruction*, *68*, 101344.

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational psychology review*, *22*, 139-154.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist*, *51*(2), 247–265.

Johnson, G. M. (2006). Synchronous and asynchronous text-based CMC in educational contexts: A review of recent research. *TechTrends : For Leaders in Education & Training*, *50*(4), 46–53.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development: ETR & D*, *48*(4), 63–85.

Jorczak, R. L. (2011). An information processing perspective on divergence and convergence in collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 207–221.

Jurafsky D, Martin J. H. (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ.

Kaiser, I., & Mayer, J. (2019). The Long-Term Benefit of Video Modeling Examples for Guided Inquiry. *Frontiers in Education*, *4*, 104.

Kalina, C., & Powell, K. C. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, *130*(2), 241–250.

Kane, M. J., Smeekens, B. A., von Bastian, C. C., Lurquin, J. H., Carruth, N. P., & Miyake, A. (2017). A combined experimental and individual-differences investigation into mind wandering during a video lecture. *Journal of Experimental Psychology. General*, *146*(11), 1649–1674.

Kawase, R., Herder, E., Papadakis, G., & Nejdl, W. (2011). In-Context Annotations for Refinding and Sharing. *Web Information Systems and Technologies*, 85–100.

Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education*, *29*(1), 89–106.

Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education*, *97*, 116–128.

Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: what is "enhanced" and how do we know? A critical literature review. *Learning, Media and Technology*, *39*(1), 6–36.

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233.

Kokoç, M., IIgaz, H., & Altun, A. (2020). Effects of sustained attention and video lecture types on learning performances. *Educational Technology Research and Development: ETR & D*, *68*(6), 3015–3039.

Koumi, J. (2006). *Designing Video and Multimedia for Open and Flexible Learning*. Routledge.

Lachner, A., Ly, K.-T., & Nückles, M. (2018). Providing Written or Oral Explanations? Differential Effects of the Modality of Explaining on Students' Conceptual Learning and Transfer. *Journal of Experimental Education*, *86*(3), 344–361.

Lam, C. N. C., & Habil, H. (2021). Enriching Student Learning Through Video-annotated Peer Feedback Activity: A Guide. *Development*, *10*(3), 46-60. https://doi.org/10.6007/ijarped/v10-i3/10712

Laurillard, D. (2009). The pedagogical challenges to collaborative technologies. *International Journal of Computer-Supported Collaborative Learning*, *4*(1), 5–20.

Lee, C., Cho, H. J., Melloch, M. R., & Levesque-Bristol, C. (2022). What worked for the engineering students to learn? Students' learning experiences through concept-point-recovery. *European Journal of Engineering Education*, 1–17.

Lee, J., Park, J. C., Jung, D., Suh, C. W., & Henning, M. A. (2021). Video learning strategies affecting achievement, learning approach, and lifelong learning in a flipped periodontology course. *Journal of Dental Education*, *85*(7), 1245-1250. https://doi.org/10.1002/jdd.12572

Lemay, D. J., & Doleck, T. (2020). Grade prediction of weekly assignments in MOOCS: mining video-viewing behavior. *Education and Information Technologies*, *25*(2), 1333–1342.

Li, N., Verma, H., Skevi, A., Zufferey, G., Blom, J., & Dillenbourg, P. (2014). Watching MOOCs together: investigating co-located MOOC study groups. *Distance Education*, *35*(2), 217–233.

Lin, Y., & Yu, Z. (2023). A meta-analysis evaluating the effectiveness of instructional video technologies. *Technology, Knowledge and Learning*, 1-35. https://doi.org/10.1007/s10758-023-09669-3

Liu, C., Wang, L., Pi, Z., & Yang, J. (2023). Interaction matters: Co-viewing facilitates learning from live video streaming for elementary students. *Education and Information Technologies*, 1–25.

Lowenthal, P., Borup, J., West, R., & Archambault, L. (2020). Thinking Beyond Zoom: Using Asynchronous Video to Maintain Connection and Engagement During the COVID-19 Pandemic. *Journal of Technology and Teacher Education*, *28*(2), 383–391.

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior*, *28*(1), 181–186.

Mamun, M. A. A., Lawrie, G., & Wright, T. (2020). Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments. *Computers & Education*, *144*, 103695. https://doi.org/10.1016/j.compedu.2019.103695

Mayer, R. E. (1996). Learners as information processors: Legacies and limitations of educational psychology's second. *Educational Psychologist*, *31*(3-4), 151–161.

Mayer, R. E. (2021). Evidence-Based Principles for How to Design Effective Instructional Videos. *Journal of Applied Research in Memory and Cognition*, *10*(2), 229–240.

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 131–152). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.009

Molnar, A. L., & Kearney, R. C. (2017). A Comparison of Cognitive Presence in Asynchronous and Synchronous Discussions in an Online Dental Hygiene Course. *Journal of Dental Hygiene: JDH / American Dental Hygienists' Association*, *91*(3), 14–21.

Morales, E., Kalir, J. H., Fleerackers, A., & Alperin, J. P. (2022). Using social annotation to construct knowledge with others: A case study across undergraduate courses. *F1000Research*, *11*, 235.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *22*(3), 149–158.

Muldner, K., Dybvig, K., Lam, R., & Chi, M. (2011). Learning by observing tutorial dialogue versus monologue collaboratively or alone. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*. https://escholarship.org/content/qt4f05w7h1/qt4f05w7h1.pdf

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.

Navarrete, E., Hoppe, A., & Ewerth, R. (2021). A review on recent advances in videobased learning research: Video features, interaction, tools, and technologies. In *Proceedings of the CIKM 2021 workshops.* Queensland, Australia http://ceur-ws.org/Vol-3052/paper7.pdf.

Nelson, B. C., & Erlandson, B. E. (2008). Managing cognitive load in educational multi-user virtual environments: reflection on design practice. *Educational Technology Research and Development: ETR & D*, *56*(5), 619–641.

Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video Improves Learning in Higher Education: A Systematic Review. *Review of Educational Research*, *91*(2), 204–236.

Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education*, *15*(1), 39–49.

Ou, C., Joyner, D. A., & Goel, A. K. (2019). Designing and developing video lessons for online learning: A seven-principle model. *Online Learning*, *23*(2), 82–104.

Oztok, M., Zingaro, D., Brett, C., & Hewitt, J. (2013). Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, *60*(1), 87–94.

Park, Y. J., & Bonk, C. J. (2007). Synchronous learning experiences: Distance and residential learners' perspectives in a blended graduate course. *Journal of Interactive Online Learning*, *6*(3), 245–264.

Peterson, A., & Roseth, C. (2015). Online Discussion Forums: The Effects of Synchronicity and Scripting. *Society for Information Technology & Teacher Education International Conference*, 436–441.

Pi, Z., Liu, C., Meng, Q., & Yang, J. (2022). Co-learner presence and praise alters the effects of learner-generated explanation on learning from video lectures. *International Journal of Educational Technology in Higher Education*, *19*(1), 58.

Pi, Z., Zhang, Y., Zhou, W., Xu, K., Chen, Y., Yang, J., & Zhao, Q. (2021). Learning by explaining to oneself and a peer enhances learners' theta and alpha oscillations while watching video lectures. *British Journal of Educational Technology*, *52*(2), 659-679. https://doi.org/10.1111/bjet.13048

Raković, M., Marzouk, Z., Liaqat, A., & Winne, P. (2020). Fine grained analysis of students' online discussion posts. *Computers & Education, 157*, 103982. https://doi.org/10.1016/j.compedu.2020.103982

Ravi, S., & Kim, J. (2007). Profiling student interactions in threaded discussions with speech act classifiers. *Frontiers in Artificial Intelligence and Applications, 158*, 357.

Reisman, A., & Enumah, L. (2020). Using Video to Highlight Curriculum-Embedded Opportunities for Student Discourse. *Journal of Teacher Education*, *71*(5), 551–567.

Retnowati, E., Ayres, P., & Sweller, J. (2017). Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *Journal of Educational Psychology*, *109*(5), 666–679.

Rich, P. J., & Hannafin, M. (2009). Video Annotation Tools: Technologies to Scaffold, Structure, and Transform Teacher Reflection. *Journal of Teacher Education*, *60*(1), 52–67.

Riessman, C. K. (2008). Narrative methods for the human sciences. Los Angeles, CA: Sage Publications.

Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science*, *36*(4), 321–350.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development: ETR & D*, *52*(1), 5–18.

Sablić, M., Mirosavljević, A., & Škugor, A. (2020). Video-Based Learning (VBL)—Past, Present and Future: an Overview of the Research Published from 2008 to 2019. *Technology, Knowledge and Learning*, *26*(4), 1061–1077.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building environments: Extending the limits of the possible in education and knowledge work. *Encyclopedia of Distributed Learning*, 269–272.

Scardamalia, M., & Bereiter, C. (2014). Smart technology for self-organizing processes. *Smart Learning Environments*, *1*(1), 1–13.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*(3), 205–221.

Schmidt, H. G., & Mamede, S. (2020). How cognitive psychology changed the face of medical education research. *Advances in Health Sciences Education: Theory and Practice*, *25*(5), 1025–1043.

Schnaubert, L., & Vogel, F. (2022). Integrating collaboration scripts, group awareness, and self-regulation in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *17*(1), 1–10.

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 375–397.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Schroeder, N. L., Chin, J., & Craig, S. D. (2020). Learner control aids learning from instructional videos with a virtual human. *Technology, Knowledge and Learning*, *25*(4), 733–751.

Schroeder, N. L., & Craig, S. D. (2017). The effect of pacing on learners' perceptions of pedagogical agents. *ACM Journal on Educational Resources in Computing*, *55*(7), 937–950.

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction*, *14*(3), 293–305.

Seaman, J. E., Allen, I. E., & Seaman, J. (2018). Grade Increase: Tracking Distance Education in the United States. *Babson Survey Research Group*. https://eric.ed.gov/?id=ED580852

Seo, K., Dodson, S., Harandi, N. M., Roberson, N., Fels, S., & Roll, I. (2021). Active learning with online video: The impact of learning context on engagement. *Computers & Education*, *165*, 1–16.

Shelton, C. C., Warren, A. E., & Archambault, L. M. (2016). Exploring the use of interactive digital storytelling video: Promoting student engagement and learning in a university hybrid course. *TechTrends : For Leaders in Education & Training*, *60*(5), 465–474.

Sherblom, J. C. (2010). The computer-mediated communication (CMC) classroom: a challenge of medium, presence, interaction, identity, and relationship. *Communication Education*, *59*(4), 497–523.

Sherer, P., & Shea, T. (2011). Using Online Video to Support Student Learning and Engagement. *College Teaching*, *59*(2), 56–59.

Stahl, G. (2003). Knowledge negotiation in asynchronous learning networks. In *Hawaii international conference on system sciences (HICSS 2003), January 6-9,2003, Hawaii, USA.*

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge (Acting with Technology)*. Penguin: The MIT Press.

Stockwell, G. (2010). Effects of multimodality in computer-mediated communication tasks. *Task-Based Language Learning and Teaching with Technology*, 83–104.

Stöhr, C., Stathakarou, N., Mueller, F., Nifakos, S., & McGrath, C. (2019). Videos as learning objects in MOOCs: A study of specialist and non-specialist participants' video activity in MOOCs: Videos as learning objects in MOOCs. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, *50*(1), 166–176.

Sun, Y., & Gao, F. (2017). Comparing the use of a social annotation tool and a threaded discussion forum to support online discussions. *The Internet and Higher Education*, *32*, 72–79.

Sun, C., Shute, V. J., Stewart, A. E. B., Beck-White, Q., Reinhardt, C. R., Zhou, G., Duran, N., & D'Mello, S. K. (2022). The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, *128*, 107120.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315–337.

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development: ETR & D*, *68*(1), 1–16.

Sweller, J. (2022). The Role of Evolutionary Psychology in Our Understanding of Human Cognition: Consequences for Cognitive Load Theory and Instructional Procedures. *Educational Psychology Review*, *34*(4), 2229–2241.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6313–6317.

Tanis, M., & Postmes, T. (2003). Social Cues and Impression Formation in CMC. *The Journal of Communication*, *53*(4), 676–693.

Vohle, F., & Reinmann, G. (2014). Social video learning and social change in German sports trainer education. *International Journal of Excellence in Education*, *6*(2), 1–11.

Walther, J. B. (1992). Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research*, *19*(1), 52–90.

Wang, A., Yu, S., Wang, M., & Chen, L. (2019). Effects of a visualization-based group awareness tool on in-service teachers' interaction behaviors and performance in a lesson study. *Interactive Learning Environments*, 1–15.

Wang J. Y., Yang C. H., Liao W. C., Yang, C., Chang, I., Sheu, B. & Ni, Y. (2022). Highly engaged video-watching pattern in asynchronous online pharmacology course in preclinical 4th-year medical students was associated with a good self-expectation, understanding, and performance. *Frontiers in Medicine*, 8, 799412. https://doi.org/10.3389/fmed.2021.799412

Watts, L. (2016). Synchronous and asynchronous communication in distance learning: A review of the literature. *Quarterly Review of Distance Education, 17*(1), 23–32

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer–supported collaborative learning. *Instructional Science*, *33*(1), 1–30.

Weisz, J. D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R. E., & Konstan, J. A. (2007). Watching together: integrating text chat with video. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 877–886.

Wong, J. K. W., Oladinrin, O. T., Ho, C. M. F., Guilbert, E., & Kam, R. (2022). Assessment of video-based e-learning in a construction measurement course. *International Journal of Construction Management*, *22*(1), 1–7.

Yoon, M., Lee, J., & Jo, I.-H. (2021). Video learning analytics: Investigating behavioral patterns and learner clusters in video-based online learning. *The Internet and Higher Education*, *50*, 100806. https://doi.org/10.1016/j.iheduc.2021.100806

Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education*, *183.* https://doi.org/10.1016/j.compedu.2022.104485. Article 104485.

Zhang, H., Miller, K. F., Sun, X., & Cortina, K. S. (2020). Wandering eyes: Eye movements during mind wandering in video lectures. *Applied Cognitive Psychology*, *34*(2), 449–464.

Zhu, E. (1996). Meaning negotiation, knowledge construction, and mentoring in a distance learning course. In: *Proceedings of selected research and development presentations at the 1996 national convention of the association for educational communications and technology*. Indeanapolis: Available from ERIC documents: ED 397 849.

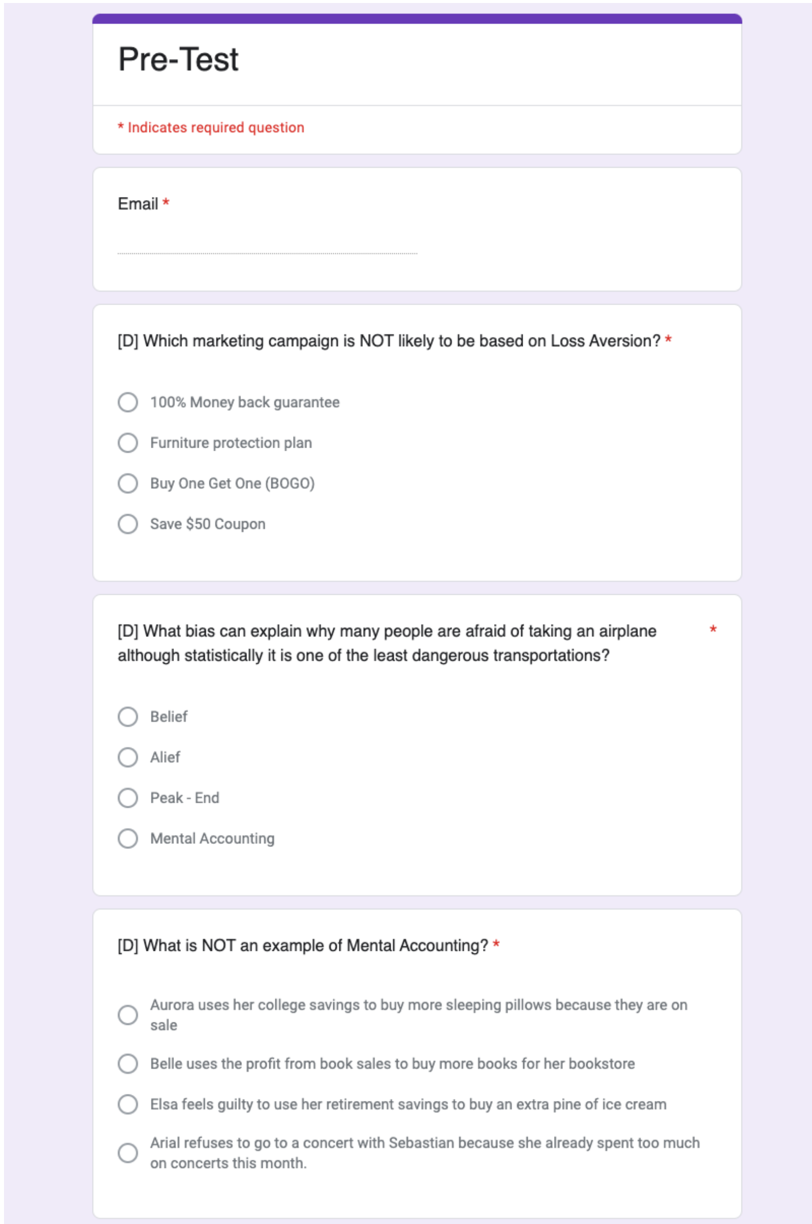
Zhu, G., Xing, W., & Popov, V. (2019). Uncovering the sequential patterns in transformative and non-transformative discourse during collaborative inquiry learning. *The Internet and Higher Education*, *41*, 51–61.

APPENDICES

APPENDIX A

ASSESSMENTS

Example of Pre-Test



Pre-Test

* Indicates required question

Email *

[D] Which marketing campaign is NOT likely to be based on Loss Aversion? *

- 100% Money back guarantee
- Furniture protection plan
- Buy One Get One (BOGO)
- Save $50 Coupon

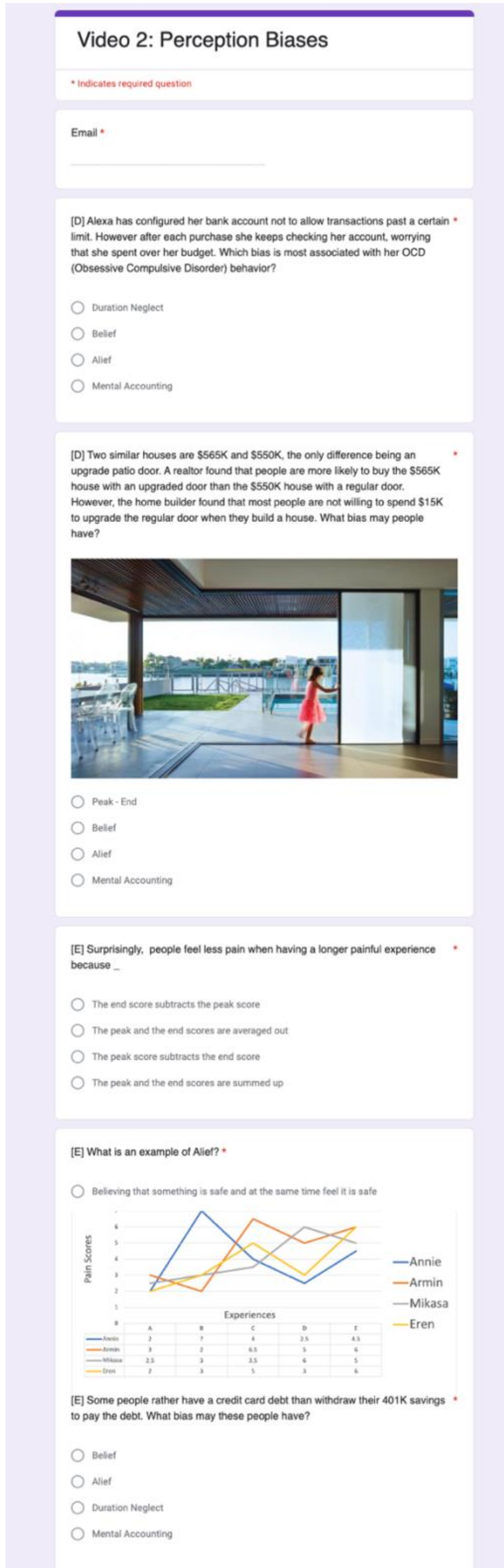
[D] What bias can explain why many people are afraid of taking an airplane although statistically it is one of the least dangerous transportations? *

- Belief
- Alief
- Peak - End
- Mental Accounting

[D] What is NOT an example of Mental Accounting? *

- Aurora uses her college savings to buy more sleeping pillows because they are on sale
- Belle uses the profit from book sales to buy more books for her bookstore
- Elsa feels guilty to use her retirement savings to buy an extra pine of ice cream
- Arial refuses to go to a concert with Sebastian because she already spent too much on concerts this month.

Example of Pro-Test: Perception Bias



Video 2: Perception Biases

* Indicates required question

Email *

[D] Alexa has configured her bank account not to allow transactions past a certain limit. However after each purchase she keeps checking her account, worrying that she spent over her budget. Which bias is most associated with her OCD (Obsessive Compulsive Disorder) behavior? *

- Duration Neglect
- Belief
- Alief
- Mental Accounting

[D] Two similar houses are $565K and $550K, the only difference being an upgrade patio door. A realtor found that people are more likely to buy the $565K house with an upgraded door than the $550K house with a regular door. However, the home builder found that most people are not willing to spend $15K to upgrade the regular door when they build a house. What bias may people have? *

- Peak - End
- Belief
- Alief
- Mental Accounting

[E] Surprisingly, people feel less pain when having a longer painful experience because _ *

- The end score subtracts the peak score
- The peak and the end scores are averaged out
- The peak score subtracts the end score
- The peak and the end scores are summed up

[E] What is an example of Alief? *

- Believing that something is safe and at the same time feel it is safe

	A	B	C	D	E
Annie	2	7	4	2.5	4.5
Armin	3	2	6.5	5	6
Mikasa	2.5	3	3.5	6	5
Eren	2	3	5	3	6

[E] Some people rather have a credit card debt than withdraw their 401K savings to pay the debt. What bias may these people have? *

- Belief
- Alief
- Duration Neglect
- Mental Accounting

Example of Pro-Test: Perception Bias

Video 1: Pricing Behavior Biases

* Indicates required question

Email *

[D] What is NOT an example of the Anchoring Effect? *

- ○ Pikachu bought Nintendo Switch at $200 and during Covid19 he sold it at $300 due to the high demand.
- ○ Piplup waited to buy a $.90 taco on Tuesday rather than pay a $1 taco on other days
- ○ Eevee set the price at $300 for his fur scarf and planed to give $50 discount if buyers bargained
- ○ Charmander bargained how much he should pay for a fire pit based on the seller's asking price

[D] Which marketing campaign is NOT likely to be based on Loss Aversion? *

○ Brow Bear's 100% Money back guarantee

○ Bath & Body Works' Semi-Annual Sale

○ Bob's Goof Proof Furniture Protection Plan

○ Safeway's Club Price

[E] Based on the video, why do people tend to like more expensive stuff? *

- ○ They associate the price with social status
- ○ They associate the price with quality
- ○ They associate the price with functionality
- ○ They associate the price with suitability

[E] A reasonable price to pay is based on a/an ____ *

- ○ Fixed Anchor
- ○ Arbitrary Anchor
- ○ Irrational Anchor
- ○ Reasonable Anchor

[E] What happens when people drink the same energy drink marked at $4 vs $2 and solve a cube puzzle? *

- ○ People feel and perform the same regardless of which can they drink
- ○ People feel less energized but perform better in the cube puzzle when they drink the $4 can
- ○ People feel more energized but perform worse in the cube puzzle when they drink the $4 can
- ○ People feel more energized and perform better in the cube puzzle when they drink the $4 can

APPENDIX B

IRB APPROVAL

EXEMPTION GRANTED

Yi-Chun Hong
Division of Educational Leadership and Innovation - Tempe
480/965-6005
shelly.hong@asu.edu

Dear Yi-Chun Hong:

On 10/26/2021 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	The effect of communication modes on online collaborative video viewing (CVV).
Investigator:	Yi-Chun Hong
IRB ID:	STUDY00014768
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Consent form-Student_CVV_V.3.pdf, Category: Consent Form; • IRB-CVV_2021025.docx, Category: IRB Protocol; • Materials for the lesson.pdf, Category: Resource list; • Recruitment Ads.pdf, Category: Recruitment Materials; • Recruitment Form.pdf, Category: Recruitment Materials;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (1) Educational settings, (2) Tests, surveys, interviews, or observation on 10/25/2021.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required.

BIOGRAPHICAL SKETCH

Growing up in Bangkok, my curiosity about life and how things work was ignited by a comic my mom gave me about world scientists. It taught me that scientists are not just individuals in white lab suits but determined, intelligent people who make a significant impact on society. Little did I know that this marked the beginning of my journey as a social researcher.

I'm deeply grateful to colleagues and professors at Chulalongkorn University, Indiana University, and Arizona State University for their unwavering support. My career has spanned the tech industry, non-profit organizations, and higher education. I started in data quality assurance and analysis and transitioned into higher education as an instructional designer and instructor. Recently, my focus has been on research in educational technology and group interaction in collaborative learning.

I've received a full scholarship from ASU's Mary Lou Fulton Teachers College (MLFTC) and travel grants to academic conferences. My dissertation was supported by the ASU Graduate & Professional Student Association's Jumpstart Grant, and a notable highlight of my Ph.D. journey was being honored with a MLFTC dissertation award.

Beyond academics, I'm passionate about data exploration, particularly through visual books that reveal unique storytelling and visualization techniques. I'm also captivated by the world of colors, their meanings, and historical significance. When I travel, I'm drawn to museums and local life, deepening my appreciation for family and the richness of life.