



# Modeling Synchronization for Detecting Collaborative Learning Process Using a Pedagogical Conversational Agent: Investigation Using Recurrent Indicators of Gaze, Language, and Facial Expression

Yugo Hayashi<sup>1</sup> 

Accepted: 14 November 2023 / Published online: 7 December 2023  
© The Author(s) 2023

## Abstract

Constructive interactions and knowledge integration activities are methods commonly used for learning; however, establishing successful coordination becomes a hurdle in computer-mediated collaborations. The development of systems to facilitate communication activities in such situations has been attempted, but models are still required for capturing learners' interactions and detecting their quality. This study explored several types of verbal and nonverbal behaviors of learners that can be implemented while designing tutoring systems to effectively capture their interaction processes in scenarios where learners engage in collaborative learning mediated by a pedagogical conversational agent (PCA). This study focused on the degree of behavior recurrence of each speaker, which is considered suitable for observing levels of effectiveness. Specifically, this study focused on three indicators—gaze synchronization, language conformance, and emotional matching through facial expression—to establish a system-based index for measuring learners' collaborative processes such as synchronization. This study experimentally examined the relationship between these indicators and the performance and process of collaborative learning among 44 learners while using PCA for facilitation. Subsequently, numerous dependent variables in the collaborative learning process were predicted using the three proposed indicators. However, no significant correlation was established between learning performance and the indicators used. These findings show that the recurrence of indicators is useful for estimating the collaborative learning process and that these indicators can be used in the development of learning support systems to trace learners' achievements in successful interactions.

**Keywords** Collaborative Learning · Pedagogical Conversational Agent · Gaze Synchronization · Lexical Alignment · Facial Expressions

---

Extended author information available on the last page of the article

## Introduction

Inspired by Vygotsky's socio-cognitive perspective, researchers have studied methods to practice learning through social interactions with others and recognize ways for learners to deepen their understanding by building on their knowledge and integrating new knowledge (Lave & Wenger, 1991; Vygotsky, 1980). This study focuses on the nature of learners' interactive processes in collaborative knowledge-building activities that have been studied in the fields of learning and cognitive science (Chi & Wylie, 2014; Miyake & Kirschner, 2014). Existing studies have indicated the enhancement of knowledge understanding through a collaborative peer (Miyake & Kirschner, 2014; Chi & Wylie, 2014), given that the externalization of an individual's knowledge can provide opportunities for elaborating on that knowledge. Studies on Learning Science have shown that constructive interactions and knowledge-integration activities in collaborative learning leverage knowledge enhancement (Scardamalia & Bereiter, 1994). Such studies have investigated explanation tasks, also known as knowledge-integration tasks, which are used in jigsaw learning (Aronson & Patnoe, 1997; Nalls & Wickerd, 2022).

However, novice learners find it difficult to achieve smooth and effective interactions, especially in computer-mediated environments where coordination becomes challenging owing to the lack of social cues. Thus, using intelligent tutoring systems such as pedagogical conversational agents (PCAs) helps facilitate collaborative processes (Hayashi, 2020). Previous studies on PCA development have primarily focused on developing systems that provide feedback to the learners for self-regulated learning and knowledge gaining (Graesser & McNamara, 2010; Heidig & Clarebout, 2011; Leelawong & Biswas, 2008; Azevedo, 2020). However, methods to model learner–learner collaboration processes for detecting learners' interactions and posing interventions to facilitate collaborative learning have not been fully analyzed. Furthermore, although studies have highlighted the usefulness of PCAs in facilitating self-regulated learning, the types of indirect facilitations useful for leading successful coordination processes in knowledge integration tasks requiring coordinative behaviors are not fully recognized. Moreover, limited studies have investigated methods to successfully capture learner–learner coordination, which can be detected through real-time learning activities using such systems. Therefore, these findings are expected to contribute to the development of tutoring systems that detect the learners' collaborative processes. Moreover, conducting real-time detection of success/failure during collaborative interaction can assist PCAs in deciding when to provide adaptive feedback and what type of feedback can lead to improvements in successful coordination.

This study aims to investigate methods of facilitating PCA development for successful learning processes and performances in knowledge integration tasks. Furthermore, this study investigates different types of indicators that can be used for capturing and modeling successful coordination processes to automate detection using PCAs, which facilitate learners' coordination processes, particularly for collaborative learning in computer-mediated environments. Such indicators include verbal and nonverbal data that are relatively easy to collect in computer-mediated environments. With such rich data, communication processes can be formalized using

methods such as natural language processing and automatic detection of learners' behavioral processes. Therefore, this study investigates the applications of verbal and nonverbal data, such as language, gaze behavior, and facial expression, collected in real-time using noncontact tools including cameras and audio devices. Furthermore, these verbal and nonverbal data are utilized to propose an index denoting the degree of synchronization of such behaviors during collaborative learning.

The remainder of this paper is organized as follows. Section 1.1 describes existing literature on explanation activities in knowledge building tasks and coordination difficulties during collaborative learning in knowledge integration tasks. Section 1.2 describes the literature of PCAs and its use in learner–learner collaboration, as well as the challenges in detecting the learning process. This section relates to the first objective of this study, which revolves around understanding how PCAs can facilitate learning activities in a knowledge integration task. Section 1.3 describes how multi-modal data can be used to detect the learning process. Section 1.4 describes the indicators of interest in collaborative activities. This section relates to the main objective of this study, which is investigating indicators for automatic detection. Section 1.5 summarizes the objectives of this study and the present hypotheses regarding the use of the indicators for capturing learning process.

## Coordination Difficulties During Collaborative Learning in Knowledge Integration Tasks

Studies have shown that different perspectives brought in by learners bring opportunities to reflect and elaborate on one's knowledge and further develop mutual understanding (van de Sande & Greeno, 2012). Empirical studies by van de Sande and Greeno (2012) verified that learners could achieve alignment of conceptual framing by adopting a schema that aligned with their common ground. Moreover, metacognition is known to play an important role when reflecting on one's knowledge while providing explanations and is a crucial factor in gaining deeper understanding (Chi et al., 1994; Chi & Wylie, 2014). Recent studies have investigated the role of social interactions in self-, co-, and shared-regulations during learning (Hadwin et al., 2018), argumentative knowledge construction (Weinberger & Fischer, 2006; Asterhan & Schwarz, 2016), and socially constructed self-regulated learning (Järvelä & Järvenoja, 2011; Azevedo et al., 2013). Learning science-based studies have used collaborative learning settings including knowledge integration tasks and jigsaw learning (Aronson & Patnoe, 1997; Nalls & Wickerd, 2022) methods studied in classrooms (Scardamalia & Bereiter, 1994; Soliman et al., 2021).

Jigsaw learning is a method in which learners acquire knowledge about different topics with separate groups. After specializing in one aspect of the topic, learners meet different group members, explain and share their acquired knowledge among each other, and develop a common understanding. Practical studies have shown that learners actively externalize their thoughts by reflecting them through explanation activities and integrating the knowledge acquired from others to build new understanding. In Chi and Wylie's (2014) ICAP (Interactive, Constructive, Active, and Passive) theory, cognitive engagement activities were classified into several modes: interactive, constrictive, active, and passive. In their theory, conflict was classified as

an interactive activity in which learners viewed different perspectives on the opinions and thoughts of others, leading to deeper discussions. Other studies have focused on key processes for developing knowledge through collaboration, such as methods to establish and achieve common ground among learners in computer-mediated learning (Chi, 2009; Rummel et al., 2009; van de Sande & Greeno, 2012).

In explanation activities, studies have confirmed the importance of successful coordination among each learner to understand others, make sense to them (Chi & Wylie, 2014), and successfully establish a common ground through conversations (Clark & Brennan, 1991; Galati & Brennan, 2021). Galati and Brennan (2021) investigated the role of modality (visual or linguistic) that speakers use to share information with their conversational partners, resulting in memory traces during referential-explanation tasks. In computer-mediated environments, communication problems become highly significant, as conversations occur in situations where the presence and awareness of others are low and limited, and a higher probability of adequate communication failure exists (Kiesler et al., 1984; Richardson et al., 2017).

Thus, a question regarding the types of collaborative processes that have been investigated and evaluated to understand effective collaborative learning activities, especially in computer-mediated collaborative environments, arises. Past studies on computer-supported collaborative learning have identified the key features of interactions and investigated evaluation methodologies for such processes (Rummel et al., 2009). Meier et al. (2007) categorized collaborative learning processes into five coding schemas: (1) communication, (2) joint information processing, (3) coordination, (4) interpersonal relationships, and (5) motivation. During communication, remarks are paraphrased to ensure understanding among the participants and alternately promote smooth exchange of messages. Joint information processing is conducted to collect information from each other to finally reach an agreement. When performing coordination, co-operating partners coordinate tasks, times, and systems. By treating each other with respect and providing opinions and perspectives, an interpersonal relationship can be formed between the participants. Finally, motivation concerns the concentration of learners on a task and the maintenance of their motivation to complete the task. Although numerous effective communication strategies for collaborative learning exist, novice learners unfamiliar with establishing joint activities in particular tasks, such as knowledge integration tasks (Hayashi, 2019a), face challenges during strategy selection. Moreover, communication through a computer-mediated environment could lack awareness, causing miscommunication (Kiesler et al., 1984; Hayashi, 2020). Therefore, such learners should necessarily achieve a successful collaborative process. However, teachers struggle to provide such prompts in class rooms with several students working simultaneously in small groups. Thus, the coordination support should be designed in a systematic manner using computer technology.

A clear understanding of the type of systems to facilitate learners' collaborative learning process is important. Therefore, the types of systems that facilitate learning processes have been further discussed.

## Facilitating Collaborative Processes Using Pedagogical Conversational Agents

In the previous section, the effectiveness of explanation activities in knowledge integration tasks for elaborative learning was discussed; however, learners face difficulties in coordination activities and require pedagogical assistance. Although computer-support systems are capable of providing such assistance, no clarity on what type of feedbacks should be used for facilitating collaborative processes exists.

To date, systems providing adaptive feedback based on human cognitive activity and their effectiveness have been studied because they relate to the development of intelligent tutoring systems (ITSs) (Anderson et al., 1995; Koedinger et al., 1997; Azevedo et al., 2013; Walker et al., 2014). For example, the early work on Cognitive Tutor was expanded to develop LISP and Geometry Tutors, aimed to teach programming languages and geometry-related topics (Anderson et al., 1995), respectively. A model-tracing approach was adopted that expresses production rules representing appropriate solutions at each step of problem solving. Recent studies on ITSs investigated learning supports using PCAs, which were based on the agent technology studied in the field of artificial intelligence (Graesser & McNamara, 2010; Heidig & Clarebout, 2011; Matsuda et al., 2015). The series of studies conducted on Betty's Brain (Biswas et al., 2005; Leelawong & Biswas, 2008; Roscoe et al., 2013) demonstrated facilitation of metacognition and self-recognition by teaching and providing knowledge to agents. A learner could use a cognitive map to solve tasks by understanding the cause–effect relationship taught to Betty. The learner could reflect on the knowledge that was taught and modify their misconceptions by teaching Betty and observing her responses. In such a learning method, learners could use the metacognitive process, which facilitated deeper understanding (Kinnebrew et al., 2014; Segedy et al., 2015).

Research using PCAs has revealed the effectiveness of methods that directly present solutions to learners and those that provide indirect feedback with no direct solutions (Shute, 2008). Indirect feedback is a form of self-regulated learning (Pintrich, 2000; Azevedo, 2020) that encourages metacognition and spontaneous learning by indirectly presenting answers to the learner (Dignath & Buttner, 2018). Researchers have accentuated the possibility of promoting learning activities by incorporating the aforementioned techniques, presenting prompts that bring awareness among learners to learn achievement goals through interactions, and presenting suggestions that evoke metacognition (Graesser et al., 2005; VanLehn et al., 2007; D'Mello et al., 2014; Azevedo, 2020). Such studies have investigated the usefulness of metacognitive feedback during learning process, but no further studies focused on the impact of such feedback in collaborative learner–learner situations.

An increasing number of studies have examined the effectiveness of collaborative suggestion presentation using PCAs. For example, learning activities that leverage the agent's physicality, such as emotions expressed by a PCA (Hayashi, 2012), and the presentation of informal feedback using modalities different from those used by learners during conversation have been studied. Furthermore, a method of using multiple PCAs to monitor learners' utterances for each role and present suggestions and facilitations that evoke metacognition has been proposed (Hayashi, 2019b). Hayashi (2019b) used metacognitive suggestion prompts as intervention prompts in a col-

laborative learner–learner setting, where the prompts were designed as indirect feedback. The results confirmed that indirect feedback enabled learners to gain deeper understanding of the explained concept. Although these studies provided interventions to learners during collaboration, the system did not have a model for detecting and evaluating sophisticated collaborative processes.

Based on existing literature and their findings, PCA can be considered as a useful technology for effectively posing indirect prompts to the learners. Hayashi (2020) demonstrated the utilization of PCAs for facilitating the collaborative learning process studied by Meier et al. (2007). This study utilizes such technologies to further investigate the influence of interventions designed for facilitating coordinative processes on collaborative learning activities. Moreover, this study investigates how the learners' reflective behavior, which are the outcomes of PCA facilitation, further enables the capture of coordinative processes. Detailed discussions on multimodal data and index are presented in the following sections.

### **Multimodal Data Modeling for Designing Support Systems Promoting Collaborative Learning**

An advantage of using computer-mediated learning environments is that the learners' data from cameras, microphones, and eye trackers can be collected and obtained at low prices. Large amounts of data on learners' verbal and nonverbal actions can be collected in real time through the devices used in learning environments. If the data are adequately processed to capture the learning process, they can be used to provide learners with adaptive feedback on ways to effectively work on collaborative tasks, which can be useful in tutoring systems such as those with PCAs.

Studies on text classification and sensing technologies have been conducted for building linguistics and behavioral pattern detectors to provide predictions on the type of conversations and interactions. For example, methods and applications were developed to analyze manually labeled speech data using natural language processing techniques (Rosé et al., 2008; Towne et al., 2017). Additionally, a study succeeded in detecting the learners' collaborative process from linguistic data and supported learning using a conversation agent (Rosé & Ferschke, 2016). These studies analyzed large scale datasets and showed improvement in detecting learners' collaborative process. Furthermore, if nonverbal cues, such as eye movements and facial expressions, were collected along with speech data, they could serve as potential predictors to increase the detection accuracy.

Martinez-Maldonado et al. (2012) researched real-time data presentation on interactive tabletops. By combining data mining results with speech and artifact manipulation, they developed an interactive dashboard that assisted teachers to monitor group activities in a multi-tabletop learning environment. Ochoa and Worsley (2016) reviewed ways to capture, process, and analyze video and audio to produce traces of actions and interactions of the actors in the learning process. They discussed analytical methods that were used to study the different modalities present in those signals and adopted in the field of multimodal learning analytics. Stewart et al. (2021) developed multimodal, team generalizable models of three key collaborative problem-solving processes: construction of shared knowledge, negotiation/coordination, and

maintaining team function. They modeled these facets in a computer-mediated video conferencing environment where the team members were free to use language, gestures, voice tone, and facial expressions to communicate.

According to the study of human interaction activities in fields such as psychology and cognitive science, determining the state of synchronized speakers is more accurate when collaborative communication occurs. Previous studies on conversational entrainment (Brennan & Clark, 1996; Galati & Brennan, 2021) and synchronization (D. C. Richardson et al., 2007; Dale, 2015) measured the degree of coordination as a quantified indicator of such interactions, which has been proven to be an effective index. Details on the index of synchronization are discussed in Sect. 1.4. These indexes are quantitative metrics derived from qualitative investigations, which detect learners' states from observations of the joint activity; these indicators may provide information about the undergoing collaborative process related to communication (Rummel et al., 2009; Hayashi, 2020). Moreover, several recent studies have investigated the use of multimodal data to capture the process of learning behavior using multiple indexes, and most of these studies have focused on verbal or nonverbal indexes of individuals (Schneider et al., 2015; Cukurova et al., 2020). Schneider et al. (2015) reviewed and identified 23 sensors more than that could be used in educational settings. The sensors included modalities such as body movements, gaze orientation, facial expressions, and physiological data including heart rate, which could be revealed as sensor-based learning platforms. If these combinations of sensor modalities that are useful in the learning domain could be determined, such sensors could be useful in understanding the learners' cognitive process during learning.

However, considering that knowledge integration is a joint activity that requires coordination between the learners, capturing learning processes at the pair level is vital. Studies should further investigate on the types of indexes that are adequate and useful for capturing successful coordination and can be collected in real time. Furthermore, the establishment of common ground in collaborative learning and non-face-to-face context, such as in computer-mediated learning in online environments, is difficult as aforementioned. Therefore, several questions arise: What type of interaction processes and conditions enable the detection of coordination quality in a learning environment? Can multimodal indexes useful for detecting the quality of collaboration be collected from camera, microphones, and eye trackers? If so, what types of processes are those multimodal indexes useful for?

In the following section, the indicators that are useful in understanding the process of collaborative learning have been explained, based on the aforementioned points.

### **Capturing Interaction from Synchronization: Recurrence and Alignment during Collaborative Activities**

As aforementioned, factors that relate to successful communication are important in collaborative learning. During an explanation activity between people with differing knowledge, the speaker typically comprehends the viewpoint of the other person and summarizes the obtained information to establish a common understanding. Words that can be understood by both participants are desirable to have an effective conversation. Communication processes, such as consensus building and the acquisition of



a common understanding, have been studied in the field of psycholinguistics (Clark & Wilkes-Gibbs, 1986; Apperly, 2018), where eye contact and synchronized activities have been established to find typical usage for promoting effective conversation. Eye contact and synchronized activities have been confirmed to generally occur with the progress in conversations (Branigan et al., 2011).

Schneider et al. (2022) conducted research on using multimodal data and developing toolkits for analyzing human behaviors. Their research included an investigation on how gaze synchronizations were correlated with collaborative process and performance. Additionally, they investigated how physiological synchrony related to collaborative learning, by comparing four different types of measures on synchrony. Relationships among higher synchrony values, physiological synchrony, and collaborative learning were established. Moreover, D'Angelo and Schneider (2021) conducted research on how shared gaze visualizations played an important role in collaborative interactions. Hayashi (2019a) conducted an analysis using synchronous indexes to investigate how the recurrences between learners mediated by PCAs influenced collaborative learning activities. The study found that collaborative processes and learning gain could be determined by the degree to which learners synchronize their gaze (gaze recurrence) and use overlapping language (lexical alignment) during their interaction. However, these two indexes were only able to detect three types of learning processes related to joint information processing and coordination. Moreover, the previous study did not conduct a comparative analysis with and without PCAs, which is an important factor that is investigated in this study. Therefore, details on what types of interactive processes that include PCAs are effective and what other potential indexes are useful for detecting interactions still remain unclear. Previous studies have combined multiple verbal and nonverbal channels to create useful indicators, but details on the index combinations for synchronization are limited and is a challenge to explore using several indexes. Thus, this study investigates emotions as extrapolated from their facial expressions along with learner's gaze behavior and language conformance.

### Gaze Recurrence/coupling

In conversations, speakers should understand similar sets of references (Schober, 1993, 2009) and use similar expressions to build a common understanding (Branigan et al., 2011). Ryskin et al. (2014) and other researchers have shown that listeners also use spatial perspective information to guide comprehension. Moreover, perspectives are modulated by basic cognitive functions including working memory and inhibition (Wardlow, 2013). Thus, synchronizing behaviors requires mental effort.

From a physical perspective, synchronization of gestures and reference to context to the same source at the same time are important (Richardson et al., 2007). Richardson et al. (2007) conducted an experiment using a rocking chair as a routine task. A pair of participants sat side by side on a rocking chair; they were experimentally manipulated to determine whether they were consciously or unconsciously attracted to the rhythm of the rocking chair and whether they were affected by visual attention. Results of the experiment, under conscious and unconscious movement conditions, indicated that the degree of visual attention affected the stability of coop-



eration between the two parties. In another study, Shockley et al. (2003) experimentally investigated the effects of visual and linguistic factors on postural coordination between two speaking people. Pairs of participants were instructed to stand face-to-face or side by side and perform a task of finding the difference in a puzzle while talking to each other or to the experimenter. Results showed that the interaction of two participants' postures was caused by communication based on verbal information and not by visual information, such as whether the other speaker could be seen. Thus, synchronized behaviors were observed even when speakers were not interacting face-to-face. However, their study did not use eye-tracking methods; hence, there is no clarity on the occurrence of gaze synchronizations.

Richardson et al. (2007) experimentally found that in reference-directed communication, the greater the proportion of synchronized gaze positions of speaker pairs, the easier it was to achieve a common understanding. In their study, they recorded the gaze of participant pairs in real-time conversations using eye trackers and analyzed the gaze patterns. They found coupling between the eye movements of the two speakers when they received the same background information prior to their conversation. Furthermore, Schneider and Pea (2014) conducted a study where two learners, each viewing an image of a neural circuit mechanism presented on a computer monitor, were tasked to explain the mechanism to the other party. The eye movements of the learner pair were measured synchronously, and the degree of gaze coincidence was calculated using the cross-recurrent analysis proposed by Richardson et al. (2007). They clarified that the greater the speaker synchronization of referencing the same parts on the screen, the better the learning performance. Therefore, the degree with which gaze information matches at a pair level during learning was established as an important index for estimating the collaborative learning process.

The existing studies showed that speaker synchronization, especially of eye movements, contributed to the success of collaborative acts. Thus, gaze recurrence during collaborative learner–learner learning, such as in explanation tasks, which are being focused on in this study, could also lead to better understanding of the content. Therefore, this study focuses on gaze recurrence as a predictor of learners' performance.

### Information Overlap in Conversation

Studies have shown that conversations are easy when speakers can align at different levels of linguistic representations (Pickering & Garrod, 2013). Speakers prime each other to speak about things in the same way, and thus they tend to think about them in the same way (Menenti et al., 2012).

In the early stages of such conversations, speakers use mutually understandable linguistic expressions to form common understanding and consensus (Clark & Brennan, 1991), which is known as language alignment or lexical alignment (Brennan & Clark, 1996). The repetition of the same words can be interpreted as a type of recurrence that can be detected from the conversational data. Linguistic alignment, when referencing a specific object, reduces ambiguity between speakers and allows speakers to proceed with conversation while reducing cognitive costs (Brennan & Clark, 1996). Moreover, speakers continue to use the conceptual pacts defined dur-

ing language alignment, even if other language expressions are available during the conversation.

In addition, Branigan et al. (2011) used a simple reference naming game to compare the likelihood of language alignment when the conversation partner was either a computer or human. They experimentally investigated how speakers align the same lexical expressions with others who could be unfamiliar with a particular phrase. They showed that when speakers believed that their conversational partner (computer confederate) had a limited vocabulary, they tended to align themselves to use the same lexical phrases to communicate. Thus, speakers tend to adapt to the other's perspective. Consequently, they tried to establish a conversation through language alignment by actively using the linguistic expressions used by the computer at the start of the task. Such language alignment can also be observed in educational situations. For example, when a teacher teaches a learner new conceptual knowledge and an ambiguous nuance arises in the expression of the new knowledge, the teacher may clarify it using the same expressions as the learner.

Thus, the aforementioned studies indicated that alignment in conversations was the byproduct of imitating each other's linguistic choices. Such conversational moves could be used in collaborative explanations; however, no studies have been conducted. Therefore, this study defines the degree to which learners can use the same linguistic expressions during conversation as a type of recurrence and investigates how the recurrence influences collaborative behaviors during explanation activities.

## Facial Expressions

The present study proposes the use of several indexes at the group level (synchronization), including capturing the degree of relationship between the learner's reactions and the collaborative process using facial expressions. Particularly, this study focuses on facial expression recognition, a computer vision technique that helps examine the emotions extracted from facial expressions as an indicator. An increasing number of studies have used facial expressions to investigate human emotional state by automated analysis (Zeng et al., 2009). Recently, studies have used the movement of facial muscles as a method for estimating an emotional state during learning activities and have developed associated software (Baltrušaitis et al., 2016). Attempts to detect emotional states using facial expressions from camera have been made, and a study has been conducted to validate them (Stöckli et al., 2018). Facial expressions can be easily collected using cameras in computers and tablets, enabling real-time data analysis. The learning situation is assumed to be open ended wherein the learners can freely discuss verbally. Therefore, exploring the applications of real-time facial expression recognition techniques in a noninvasive situation is crucial.

The identification of emotional states from facial expressions has been a long-standing issue in emotion recognition research. Paul Ekman's definition of universal facial expressions for six emotions (disgust, fear, anger, sadness, surprise, and happiness) and his proposal of a mapping method based on facial expressions have further advanced the research on emotion recognition (Engelmann & Pogosyan, 2013). The proposed method has made rapid progress since the late 1990s, and a method for labeling Ekman's basic emotions from multiple patterns of facial expression muscles

using still images and audio-visual signals has been established as the Facial Action Coding System (FACS). The method was used to successfully identify the emotional state of a speaker (Skiendziel et al., 2019). The validity of labeling emotional states using the proposed method has been confirmed by researchers such as Bartlett et al. (2005).

Considering these perspectives, this study focuses on the six basic emotions that are useful and accurate for detection in real-time computer-mediated tasks. In collaborative learning activity using PCA, such as in this study, the basic emotions are interpreted as (1) joy: when an understanding is reached or a consensus is formed during discussion, and when a sense of accomplishment or satisfaction is achieved in a task; (2) sadness: when the concept to be explained is difficult to understand; (3) anger: when frustration or dissatisfaction builds up due to poor explanations; (4) disgust: when communication is not going well and a negative feelings such as a discomfort about the explanation arises; (5) surprise: when learners discover something new or unexpected about the content from the other person's explanation; and (6) fear: when the task is not progressing well, or when the submission time for the task is approaching. In the field of learning science, studies on other types of emotions have been conducted. For example, Graesser et al. (2005) showed that learning had a positive correlation with confusion, involvement, and psychological flow; negative correlation with boredom; and no correlation with any other emotions. However, determining the type of social emotions for learning is still debatable. Additionally, to the author's knowledge, software for noninvasively detecting such social emotions in real time is not commonly used.

Conversely, software that can noninvasively, automatically, and in real time calculate emotional states based on the FACS model have been developed. Additionally, studies have used face reading tools that detect basic emotions as a method to capture learners' activities during learning tasks. For example, Face Reader (van Kuilenburg et al., 2005) utilizes FACS to distinguish between the six basic emotional states with 89% accuracy. Moridis and Economides (2012) used this software to detect the basic emotional states of learners in an e-learning environment, and conversational agents responded in an empathic manner according to those emotions. Furthermore, Alkilani & Nusir (2022) conducted a study wherein a conversational agent was able to detect the learner's basic emotions using Face Reader and respond empathically. Furthermore, they developed a system that could detect learners who were likely to drop credits by using the emotions detected from their facial images in online examination situations. Further investigations of basic emotions that were likely to occur in each state in the ICAP of Chi (2009) were conducted and quantitative examinations were performed (Cai et al., 2020).

Based on the aforementioned literature, this study expands this line of research by focusing on the degree of alignment of learners' emotional states detected using FACS. When learners work collaboratively, they share the problems they encounter; therefore, they may have similar facial reactions (for example, a negative attitude or expression when reaching a problem-solving impasse or celebration at a mutual agreement or success). Therefore, this study focuses on the synchronization of learners' emotional states extracted from their facial reactions, and predicts similar reactions from the learners when they are successfully coordinating with each other.

The next section summarizes the objective of this study and provides the hypothesis related to each goal.

## Objective and Hypotheses

As discussed in the Sect. 1.1, novice learners in a computer-mediated environment face coordination difficulties while working on explanation activities during collaborative learning. As discussed in Sect. 1.2, PCAs facilitate self-regulation and better coordination among the learners. Moreover, to develop automatic detection of successful coordination by PCAs during learner–learner collaborative activity, investigating the type of indexes that can be used for detection while using available technology is important. Therefore, this study aims to investigate (1) methods that can lead to successful learning processes and knowledge integration performances in knowledge integration tasks facilitated by PCAs. In terms of the second objective, which is the main focus of this study, the investigation focuses on (2) the type of indicators that should be used for capturing a successful coordination process for automatic detection while using PCAs, which facilitate the learners coordination process. This study investigates the applications of verbal and nonverbal data collected in real time using noncontact tools, such as cameras and audio devices.

To attain the first objective, this study focused on collaborative learning of learner–learner dyads in explanation activities during a jigsaw-like knowledge building task, wherein coordination between the learners was essential for accomplishing the task. A modified PCA from previous studies (Hayashi, 2012, 2019a, b, 2020) was used to mediate the learners' interactions and provide suggestions to two learners for improving their coordination. The interventions provided by the PCA reminded the learners of their goals and motivated them to focus on their task, which was to coordinate with each other using explanations and turn-taking to develop mutual understanding about the unknown texts presented on their screen. Moreover, the PCA facilitation included not only metacognitive content but also facilitation related to social content, such as methods that motivate them to take turns and coordinate. Therefore, this study predicted that PCA facilitations led to improved synchronization between the learners and could be used as a good indicator for capturing the quality of the collaborative learning process and performance. Therefore, this study examined whether facilitation using a PCA promoted the process and performance of collaborative learning. The hypothesis (H1) for this examination is stated as follows.

**H1-1** Learners who receive facilitation through PCA will experience better-quality collaborative learning than others who do not.

**H1-2** Learners who receive facilitation through PCA will achieve better performance in collaborative learning than others who do not.

As discussed in Sect. 1.4, several studies have used sensing technologies and multiple types of quantitative data for their analysis. However, limited work has been done on multimodal indexes designed to enable the detection of collaborative processes at the pair level, and no established methods exist that can be applied to knowl-

edge explanation activities in computer-mediated learning settings. Previous studies on detection of learner collaboration in tutoring systems primarily focused on the individual level of learners' activities; therefore, further development of adequate indexes at the pair level is important. In past psychological research, alignment and synchronized behavior between speakers have been established to lead effective communication, such as in coordination tasks (Richardson et al., 2007). Such studies have focused on synchronization, which is an index that can be collected at the group level from dyads. Thus, this study focused on the recurrence degree of dyadic collaborative interactions in a jigsaw-like knowledge integration task, where coordination is required to accomplish the task. This study aimed to use linguistic and nonverbal data collected from contactless devices to determine the degree of synchronization.

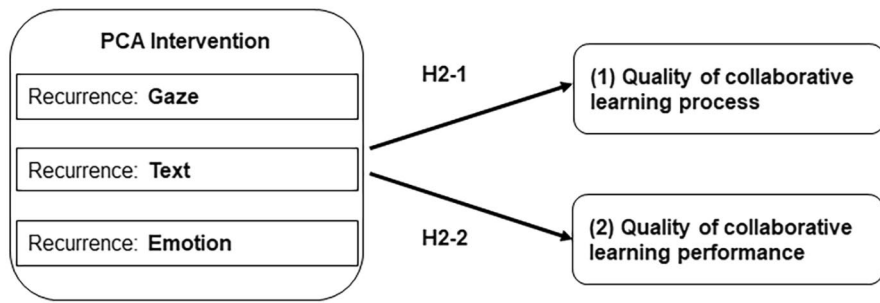
For the second objective, three recurrence indicators, gaze, text, and emotion, were used to reflect the collaborative process, assuming that the PCAs intervening during the collaborative process could reflect directly on these indicators. Learners who received facilitation on collaborative learning methods could easily synchronize their actions by behaving according to the content. Therefore, the hypothesis was that the recurrent indicators were strong predictors of quality learning during collaboration with PCAs. Moreover, if the learners followed the interventions from the PCA and synchronization occurred, learners were expected to provide better explanations of the concept and deepen their understanding. Therefore, recurrent indicators could also be strong predictors for learning performance. However, learning performance was evaluated after the task and was not directly linked with the learning process; different cognitive steps could be required. Therefore, unlike with the learning process, the three predictors could not be strong predictors of learning performance. Thus, the three recurrent indicators were predicted to affect the process and performance of collaborative learning. Hence, the second hypothesis (H2) is as follows.

**H2-1** When using a PCA, recurrence of the specified indicators can predict an improvement in the quality of the collaborative learning process.

**H2-2** When using a PCA, recurrence of the specified indicators can predict an improvement in the quality of the collaborative learning performance.

To examine the hypothesis (Fig. 1), a laboratory-based study was conducted using a simple knowledge integration explanation task (Hayashi, 2020), where the learner pairs performed explanatory activities and integrated conceptual knowledge. A laboratory-based study focusing on peer learner–learner collaboration, wherein the learners conversed with each other throughout the main task, was conducted. Hayashi (2019b) used a text-based chat and a PCA that automatically detected the utterances to provide feedback. Contrarily, free settings were used in this study, and learner conversations were allowed. An important part of this arrangement was that the learners were required to look at the computer screen the entire time, enabling the recording of the entire learner's face and gaze movements during the task.

The following section describes the methodology followed for the experiment and the details of the indicators used in the experiment.



**Fig. 1** Main hypothesis of this study

## Methods

Experimental investigations were performed on the collaborative learning in a learner–learner explanation activity, which comprised jigsaw-based learning in a non-face-to-face setting. The settings were chosen to investigate the aforementioned points in situations where learners may face discrepancies and awareness of each other is low, as in online distant learning. A jigsaw-like paradigm was designed based on the experimental paradigm used by Hayashi (2020). To summarize, learners were tasked to verbally explain two different types of knowledge to each other, integrate the two, and create an explanation of the entire content. In this task, learners were expected to coordinate with each other by following these steps: (1) each learner read the text on their screen to their partner, (2) they developed a common understanding of the two concepts by asking questions, and (3) they developed a new abstract understanding of the topic (“how human information processing works”) using the two concepts. Particularly, either one of the different types of knowledge was shown to one learner, and they were asked to teach their partner about it. The two texts were presented separately on the computer screen using the same layout for both, except that one of the texts was covered (not visible); thus, the learners had to teach their partner about their text to accomplish the task. Learners only communicated by voice and sat such that they could not see each other or the other’s text.

## Participants and Conditions

Forty-four Japanese university students (female: 24, male: 20, average age: 20.79, standard deviation SD: 1.93) majoring in psychology participated in this study. They were recruited via an experimental pool operated by the author’s department and registered for the experiment on a first-come-first-served basis. They received bonus course credit upon participation in the experiment, which was conducted in dyads. Twenty-two participants (11 dyads) were assigned to the control group without a PCA, while the remaining twenty-two (11 dyads) were assigned to the intervention group in which a PCA provided intervention. All participants (“learners”) were grouped in a random order within the same gender. A previous study showed that dyads were more active in communication when in groups of the same gender (Rohr-

beck et al., 2003); thus, this study followed the same procedure. The learners were not provided with the details of the experiment, such as interventions, beforehand and only received debriefing after the experiment. The experiment and this study were conducted after receiving approval from the ethical review committee of the author's university.

## Procedure

Upon the participants' arrival at the experiment site, the experimenter thanked them for their participation and introduced them to their partners. The experimenter provided instructions about the task to be performed, which was a jigsaw-like explanation task that combined two different technical concepts (topics in cognitive science) to explain a general question of how human mental process works, as discussed in Sect. 2.3. Before the main task commenced, the learners completed a free recall test covering the concepts that would be referred to during the main task to check their knowledge of the concepts. Subsequently, they completed the main explanation task in approximately 10 min. After the main task, they completed another free recall test. Upon completing the entire experiment, they were debriefed.

## Task

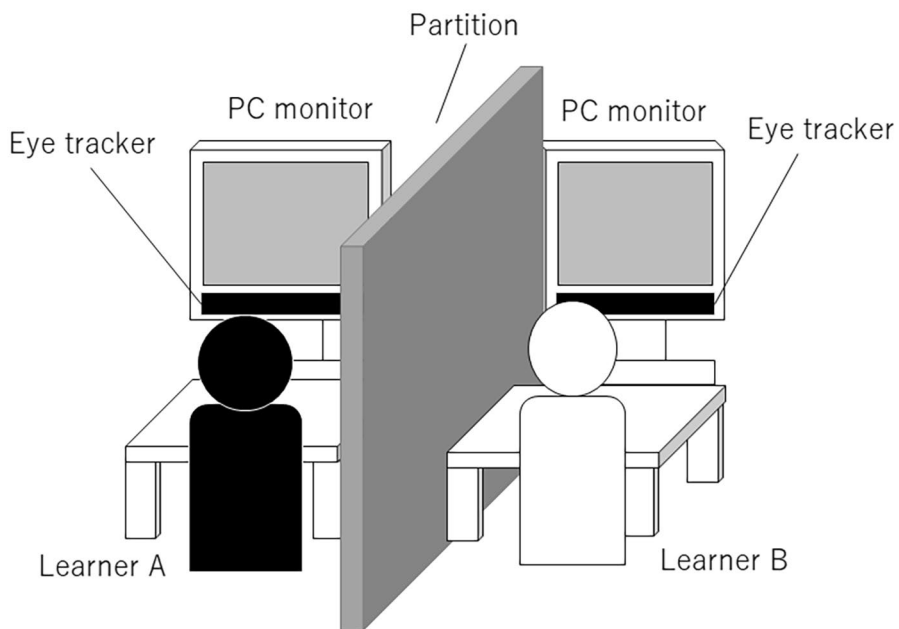
The task required participants to explain a cognitive science topic (i.e., human information processing related to language perception) using two specialized concepts (i.e., “top-down processing” and “bottom-up processing”). In this study, the knowledge integration task was conducted according to the procedure used by Hayashi (2019a). In this paradigm, learners were required to orally explain contents (concepts) to their partner who had no knowledge about the content. Learners took turns to explain their concepts and then integrated them. Throughout the activity, the learners were expected to develop an abstract understanding of the two different concepts and construct higher abstract knowledge about the learning material; here, human mental process was to be understood based on the knowledge explained by the learners. They reached a common understanding by explaining the content to their partner, who could then use the acquired knowledge to further explain the entire concept. First, each learner was asked to read the text on either “top-down processing” (concept A) or “bottom-up processing” (concept B).

- Top-down processing (concept A).
  - A way of thinking that allows you to understand what you are seeing and hearing based on existing knowledge and ideas that come to mind.
  - Information processing in the language in mind depends on factors such as the person's knowledge, expectations, and attitude.
  - Top-down processing, also called concept-driven processing, is a processing method that largely depends on human memory. It is driven by concepts and theories at higher levels, and the input data are based on expectations, hypotheses, and observations.

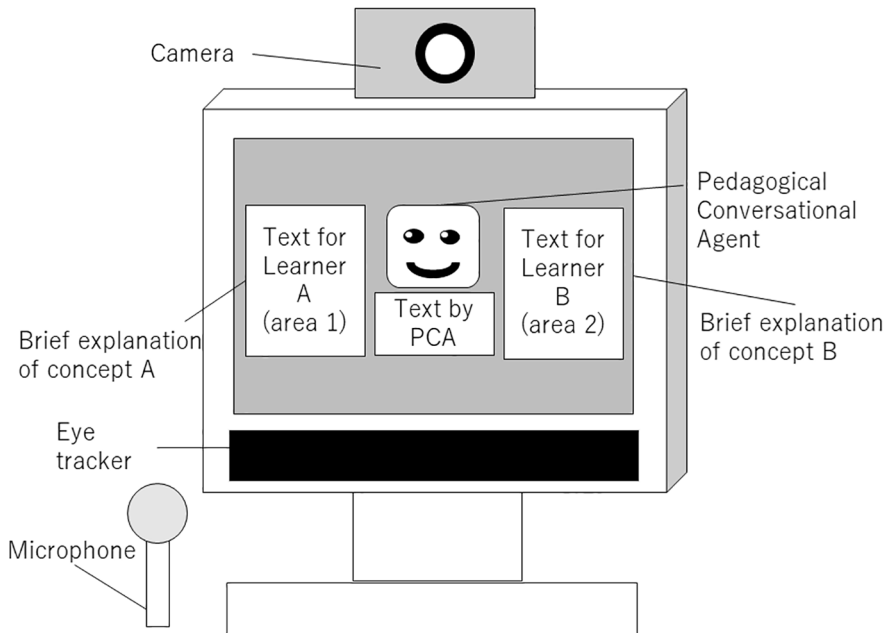


- Bottom-up processing (concept B).
  - A way of thinking that analyzes the constituent elements of the sentence you are looking at and the language you are listening to and determines the meaning from the analyzed characteristics.
  - Information processing in the mind depends on the physical, spatial, and temporal arrangement of the object.
  - Bottom-up processing, also called data-driven processing, is driven by the feeling of receiving input from the outside world, and it performs processing by finding a schema (knowledge) to handle and process it.

In this study, the concept that the learner read was referred to as “self-concept,” and the concept that their partner read was referred to as “other-concept.” After the learners finished reading their text, they explained the concepts to each other. During their explanations, the PCA intervened in their conversations and prompted them. To experimentally construct a situation in which learners cannot know both concepts, the learners were separated and the was presented in a specific manner; the participant pairs were seated at desks separated by a partition (Fig. 2), and the experimental task was divided into two phases. In the first phase, as a pre-task, they individually read their self-concept within 5 min. In the second phase, they talked about their self-concept for 10 min as in the main task. During the main task, the learners saw a displayed text summarizing the self-concept and other-concept masked with a blurred image next to each other (Fig. 3).



**Fig. 2** Experimental setting



**Fig. 3** Image of a learner's screen. Both screens have the same layout

During the task execution time of 10 min, one message per minute was sent from the PCA regarding the content of the learner's explanation when the conversation stopped or when there was a momentary pause. The details of the designed PCA are described in Sect. 2.4.

The face of the PCA and its chat text were displayed in the center of the screen. The self-concept was displayed on either the left or right side of the screen in the form of a summary, while the other-concept was displayed as a blurred image so that the learner could not simply read the information on their partner's screen and thereby gain understanding. The layout of the computer screen and contents were the same between the two learners except for one change in terms of the blurred area. The learners' first task was to read and explain the content that was presented and could be seen only by one of them. This manipulation was to create a situation in which learners did not know each other's concept and had to coordinate with each other, explain, and ask questions about what was written in the blurred area to achieve the task objective.

In addition, before the beginning of the experiment, the learners were instructed about the layout of the screen and the process of the task. Particularly, the explanations written in each area were read out in order after the task begun, and the explanation activities were performed while taking turns. One of the key points was that the learners had to receive an explanation from their partner to learn the other-concept. In the author's previous research (Hayashi, 2020), when a learner read aloud a sentence from their self-concept, the other learner looked at the blurred area to comprehend what was written there. As aforementioned, during explanation, learners could exter-

nalize their thoughts and request for further details to deepen their thoughts; thus, metacognition was expected to be used during reinterpretation and abstraction of their thoughts.

For example, Learner A could start looking at the other-concept area to request an explanation about what was written in that area, such as “I want to know about your concept. What is written in this right area? I cannot read it because it is blurred out. Can you start reading the first sentence, please?” Then, if Learner B accepts this, they look at the self-concept area and start reading out the explanation. As the interaction progresses, their gazes synchronize, and they both look at the same area; this was calculated as high recurrence for the gaze index. Moreover, learners could use the same phrases and technical words when rephrasing (e.g., “Did you say that top-down process uses knowledge and experience from the past?”), and therefore similar words could be used at the lexical level. Moreover, if the learners were successful in coordinating and confronting a shared problem during the task, such as a confusion, or succeeded in obtaining the entire explanation, they could feel the same way, thus showing similar emotional states.

To test the hypothesis, the eye movements and audio data from each pair of learners collected during the main task of the experiment were analyzed. Two eye trackers (Tobii X2-30) of 30 Hz were used to collect simultaneous measurements of eye-movement data. The two eye trackers did not require a fixed head position; thus, the learners were free to move their head and talk. The eye trackers were attached to the computer monitor, and the sensors installed in the devices captured the learners’ eyes. The details of the analytical method are described in Sect. 2.6.1. Furthermore, to analyze the learning process of the learners’ conversations, the audio data of the conversation for the entire duration of the task were collected and then transcribed.

## Experimental System

This study used a PCA that provided informal feedback to facilitate learners’ self-regulation and metacognition. The PCA was situated between the two collaborating students and mediated the collaboration between them. This study improved and used a simplified version of an online chat platform for collaborative learning with PCAs, which was developed by the author in a previous work (Hayashi, 2019b). In this system, the content that was deemed inappropriate in the experiments conducted in the previous research (i.e., prompts that the learner did not respond to) and the facilitation method (i.e., presentation frequency and variation) were updated. To investigate how synchronization occurs using the recurrence indexes, this study allowed learners to only interact by speech and did not use any text-based chat. This condition was important because it was necessary to collect the learner’s stable gaze and facial expressions to examine the degree of synchronization. For detecting the speech of the learners, this study used the Wizard-of-Oz (WOZ) method. Speech recognition was also a potential option; however, the author decided to use the WOZ method for inputting the speech into the system to provide a timely and accurate response. The experimenter listened and inputted the learners’ conversation as text into the PCA installed in each terminal. Then, the PCA responded by generating voice. In addition, the PCA intervened in the learners’ conversation at intervals of approximately

1 min and sent a facilitation message when the conversation was interrupted. When performing facilitation by PCA, the displayed mouth and upper body were designed to move, and this movement adopted a frame-by-frame advancement method of multiple image frames at 100 ms intervals. For facilitation, this study used remarks that encouraged reflection on achievement of goals (Azevedo & Cromley, 2004), encouraged metacognition, and promoted motivation (Hayashi, 2012). The following five types of facilitation were prepared.

- Type A: Facilitation reflecting on the purpose of the task and the significance of achieving the objective. (e.g., “Keep in mind that the goal of this task is to create an entire explanation using the two concepts.”)
- Type B: Facilitation promoting metacognition. (e.g., “You are using important words written in the text. Try to reconsider why the wording is important and explained in the text.”)
- Type C: Facilitation encouraging the other person to pay attention to the knowledge content being read. (e.g., “It is important to first listen carefully to your partners’ explanations so that you can explain it by yourself and achieve the goal of this task.”)
- Type D: Facilitation that motivates one person to speak. (e.g., “Keep going. Explain to each other the text the other does not know.”)
- Type E: Facilitation that encourages focus on the task. (e.g., “It seems that you two are not using words from the text. Try using the words from the text.”)

Types A and E were prompts related to encouraging reflection on achievement of goals, Types B and C were related to encouraging metacognition, and Type D was a motivating prompt. Type B generated reflective prompts based on the detection of important keywords that were used related to the concepts (i.e., “schema” or “module.”), while Type E generated reflective prompts when no such important keywords were used after 5 min of the conversation. Suggestions such as advice to use the words from the text and instructions to reconsider the text were provided in an abstract manner. The prompts were sent at 1 min intervals. The rule for WOZ feedback was (1) if the participants were using important keywords such as “schema,” “module,” or “transfer,” then feedback was presented (Type B), and (2) if such keywords were not used, then other prompts were presented randomly (Types A, C, D, or E). The same experimenter played the role for WOZ. These facilitations by PCAs were an indirect facilitation for self-regulation and indirectly gave orders that were biased on recurrence, such as ordering learners to look at the same area (gaze recurrence), use the same words (lexical alignment), or react in the same way (facial recurrence).

## Measures

This section describes the dependent prediction variables used in this study for the process and performance of collaborative learning.

## Performance

The learners were asked to explain in detail what they knew about each of the two concepts used in the experiment (self-concept and other-concept) on an answer sheet. After the experiment, the description of the content given as answers in the pre-test/post-test were divided into a description of the self-concept, an explanation of the other-concept, and an integrated description of both the self- and other-concepts. These were respectively scored based on the following scoring criteria.

The scoring criteria and some of the actual response examples are given below.

- Incorrect answer/No answer: 0 points.
- “Ability to understand words. It is natural for humans to understand sentences, phrases, and contexts, but it is exceedingly difficult for machines.”
- The answer can be interpreted as correct but does not use appropriate technical terms described in the learning text. The answers are written based on naive individual inferences: 1 point.
- “In the top-down model, which is one of the theories about language perception, inferences based on the current information were made. Bottom-up processing, on the contrary, picks up information and processes the information based on it.”
- The answer is correctly described using appropriate terms used in the learning text. The answers are a summary or repetition of what was described in the text: 2 points.
- “When perceiving a language, there are two types of processing methods: top-down processing and bottom-up processing. Bottom-up processing analyses information obtained from the outside world, divides it into elements, finds the generality in it, and processes it.”
- The answer is correctly described using appropriate terms described in the learning text and uses further interpretations and inferences, such as analogies: 3 points.
- “Top-down processing is also called concept-driven processing and is a function where processing depends on expectations and hypotheses that depend on human memory. Linguistic understanding of a sentence requires elements to be connected and simultaneously processed and understood as a single coherent statement by fully utilizing previously encountered concepts and experiences according to the situation. Bottom-up processing is, for example, understanding a sentence using a language, picking up the elements contained in the sentence, perceiving each one, and processing it in the mind. Language is perceived by bottom-up processing and top-down processing. It may be difficult to perform these two processes in parallel, but the processing method may change depending on whether a new or familiar word was used.”

To score responses according to the above description, two coders coded the experimental responses after discussion using a rubric. Thus, the coding agreement rate was 78.6%. The intra-class correlation was calculated and found to exceed 75%. The author reviewed all inconsistent items and classified them accordingly. For items with a discrepancy, the average value of the rating scores was calculated and exam-

ined using the rated score as an index. The gain score was calculated using the pre-test and post-test scores, as follows:

$$\text{gain} = (\text{post test score} - \text{pre test score}) / (\text{maxscore} - \text{pre test score}) \quad (1)$$

## Quality of Collaboration

The utterance data collected were transcribed, and an utterance analysis was performed. In the speech analysis, the rating scale of the collaborative learning process proposed by Meier et al. (2007) was used. The scale was classified into five categories: (1) communication, (2) joint information processing, (3) coordination, (4) interpersonal relationship, and (5) motivation. It consisted of nine sub-item groups, which was organized as follows.

### (1) Communication.

(1-a) Sustaining mutual understanding — Did the learners have a conversation that provided a common understanding of the concept? (e.g., “So let me clarify, what you said about a schema; this was about the knowledge acquired from experience and used for top-down processing”)

(1-b) Dialog management — Was the conversation going well? (e.g., “So, I think I understood well what you meant. Thank you. Let us take turns and I will start next.”)

### (2) Joint information processing.

(2-a) Information pooling — Did the learners attempt to collect information about each other’s concepts? (e.g., “So now that we have both read the texts, which we are supposed to explain, is there anything left that I should know?”)

(2-b) Reaching consensus — Was there a final consensus on the concept? (e.g., “I think we are good now, right? Should we summarize these further now and start to think about how these could be used for further explanations?”)

### (3) Coordination.

(3-a) Task division — Did the learners proceed with the tasks sequentially? (e.g., “Let us divide the task in two parts. Then, why don’t we each explain our concept taking turns and ask each other some questions.”)

(3-b) Time management — Was time managed properly? (e.g., “I think we should start aggregating the things in the last three minutes because time is limited.”)

(3-c) Technical coordination — Did they understand the technical content? [not used]

### (4) Interpersonal relationship.

(4-a) Reciprocal interaction — Was the information provided in a reciprocal manner? (e.g., “I think I am talking too much and using the time, so why don’t you use the rest for explaining your concept to me.”)

(5) Motivation.

(5-a) Individual task orientation — Was there enthusiasm for the task? (e.g., “This is interesting and the more I read it, the clearer it gets, and I wonder how it is described on your side. Please let me know when you are ready.”)

Based on the above classification scale, the utterances of each learner pair were rated by two experts. The coders discussed the rubric before starting to code. The utterances were scored on a 5-point scale (–2: very not applicable – 2: very applicable). The concordance rate of the evaluator’s utterance ratings was calculated using Cronbach’s  $\alpha$ , which was equal to 0.71. The observed value was higher than the standard value of 0.7. For items with a discrepancy, the average value of the rating scores was calculated and examined using the rated score as an index. The intra-class correlation was calculated and was found to exceed 74% for all measures. Similar to the study by Schneider and Pea (2014), in this study, the technical coordination was excluded from the analysis because it was not examined in the considered environment (no discussion in technical operation methods was available).

## Index Used for Prediction: Predictors

This study focused on collaborative learning based on knowledge integration in which learners had to successfully coordinate with each other and establish a common ground. To automatically detect such learning process in real time, it was necessary to model the learners’ coordination process. Three types of synchronizations—gaze recurrence, language conformance, and emotional matching—were studied as valuable indexes for capturing the level of success in coordination.

## Gaze Recurrence: Analysis of Gaze Recurrence

Gaze synchronization is an important activity for successful communication. Based on the findings of a previous study by Schneider and Pea (2014), this study analyzed instances where the learner pairs were looking at the same screen area simultaneously. Schneider and Pea (2014) considered a deviation of  $\pm 2000$  ms and determined that the greater the degree to which both paid attention to the same area, the more successful was the process of understanding the other’s perspective. The analysis in this study was performed according to similar criteria. In their past study, Richardson et al. (2007) examined the direction of the gaze of speakers and listeners on the same visual scene and found that the gaze shifted towards the same area after referential utterances. They found that the listener’s eye movements most closely matched the speaker’s eye movements with a delay of 2000 ms. Contrarily, the current study used simpler stimulus and informed the learners where the stimulus would be presented. Therefore, the time lag of 2000 ms was considered as sufficient for the

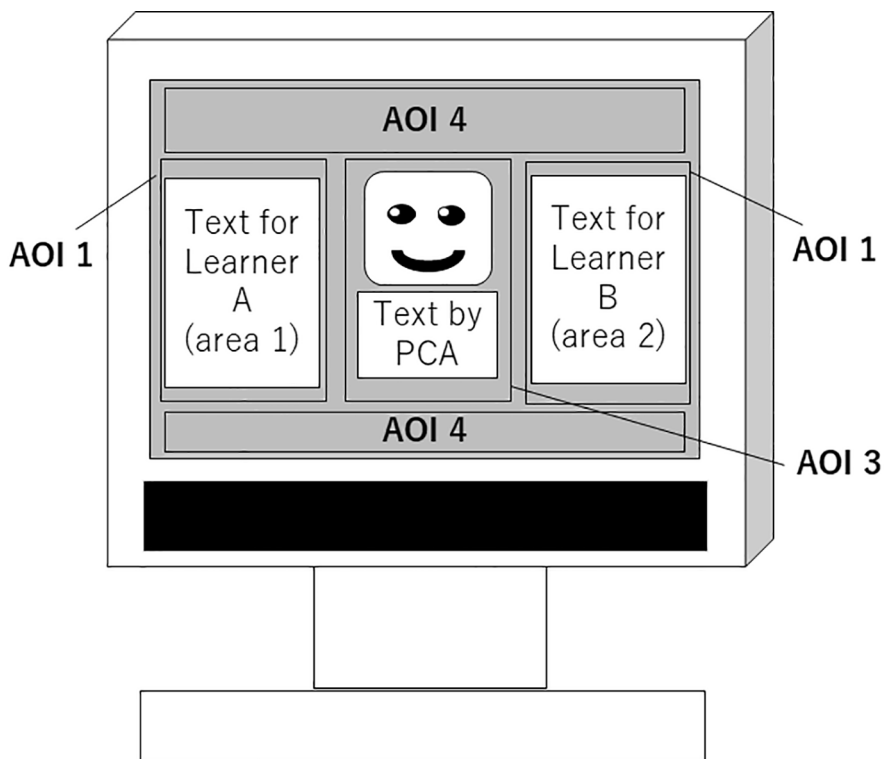


required analysis. For the analysis, the screen area was divided into the following four areas (Fig. 4).

- Area of interest 1 (AOI 1): Area on the left where a concept was presented.
- Area of interest 2 (AOI 2): Area on the right where a concept was presented.
- Area of interest 3 (AOI 3): The central area where the face of the PCA was presented.
- Area of interest 4 (AOI 4): Areas other than specified above.

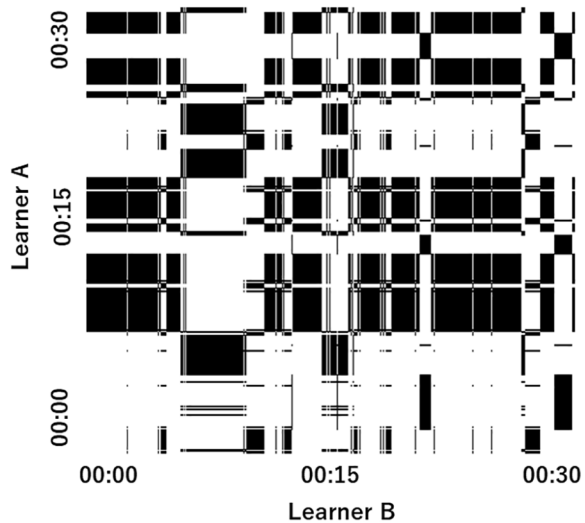
For each screen area, the number of gaze dwelling moments for each learner was counted, and the rate of matching gaze between the learners was calculated based on the method described by D. C. Richardson et al. (2007).

This method was used to analyze whether learners were looking at the same area simultaneously. Specifically, the recurrent phi was used calculate the percentage of Learners A and B that matched on the time axis,  $k$ . The higher the concordance rate of  $(k; k)$ , the larger the value of phi. Figure 5 shows an example of the gaze recurrent analysis, and the analysis in the figure considers the time lag, which follows the same procedure as previous studies (Schneider & Pea, 2014; D. C. Richardson et al., 2007). In this figure, the vertical axis shows the timeline of Learner A, and the



**Fig. 4** Division of screen area as used in this study (AOI: Area of Interest)

**Fig. 5** Example of gaze recurrence analysis. Dark blocks indicate mutually aligned gazes



horizontal axis shows the timeline of Learner B. Pixels are marked every 200 ms, and the black-filled areas indicate the gaze synchronization of Learners A and B on the time axis,  $k$ . Meanwhile, the white areas indicate that the gazes of the learner pair did not match. In the analysis,  $\phi$  was calculated for each participant for the region considering the deviation of  $\pm 2000$  ms in the gaze coincidence between Learners A and B in the figure and used as an index for prediction. The mean (standard deviation) of  $\phi$  was 0.229 (0.055).

### Language Conformance: Adopting an Epistemic Network

To examine the degree of language alignment between the learner pairs, the vocabulary used by the learners during the conversation was analyzed. It was assumed that a greater degree of matched vocabulary used by the learners during the task indicated better language alignment. It may be argued that language alignment is not exactly an index of synchronization. However, as mentioned in Sect. 1.2.2, this study defines the repetition of language use as a type of recurrence that can be detected from conversational data and useful for evaluating learning behavior. In the analysis, (1) the words used by the learners during the task were extracted and (2) the degree to which the words were used more than once by the pair was calculated.

In (1), a list of words was created for each learner, and each of their usage frequencies were identified. To achieve this, morphological analysis was performed using the RMcCab package in R, and nouns that appeared more than once were extracted. Furthermore, words not related to the study (e.g., “I” and “you”) were manually deleted from the list. Next, a dictionary list of keywords extracted for each learner pair was created. Examples of words contained in the dictionary lists included “processing,” “perception,” “knowledge,” “experience,” “input,” “instinct,” “element,” “whole,” “past,” and “meaning.” In (2), the match rate of the words that each learner said was

calculated using the dictionary list created for each learner pair. The author manually edited some of the words that were semantically similar.

The match rate  $k$  was calculated using Eq. (2).

$$k = l / n \quad (2)$$

where  $n$  indicates the total number of words in each of the dictionary lists of Learners A and B, and  $l$  indicates the number of words that were matched in both. Here, the interpretation suggested that the closer the value of  $k$  is to 1, the higher the language alignment between the learners. The value of  $k$  was calculated for all learner pairs, and the language alignment value was calculated for each individual.

### Facial Pattern Matching

This study used a method for detecting emotional states from facial expressions collected using a video camera (Sony, HDR-CX680). The learners' chairs were adjusted so that facial images were captured, and their facial expressions were recorded during the main explanation task. For the facial expression analysis, this study used Face Reader (<https://www.noldus.com/>) to evaluate the emotional states of the learners during the interactions. Face Reader was used because it processes video images pre-recorded by video cameras, and it was shown to be reliable in the author's preliminary study on mapping emotional states and facial expressions (Hayashi, 2019c). As aforementioned in Sect. 1.4.3, this system could classify expressions as one of six emotional categories: joy, anger, sadness, surprise, fear, and disgust (van Kuilenburg et al., 2005). The tool recognized fine-grained facial features based on minimal muscular movements described by FACS. The system used an active appearance model for creating a model of the facial expressions for classification, where the shape of the face was defined by a shape vector that contained the coordinates of the landmark points (van Kuilenburg et al., 2005). Prior to the video recording of learner's facial expressions in the experiment, the experimenter collected six emotional states of each learner to calibrate the emotional baseline of each individual. The experimenter asked the learners to express each emotional type before the task started. The fundamental facial expressions collected were used as the neutral baseline for calculating each emotional state using the software.

The calculations used for the recurrence index for emotional states were the same as that used for gaze recurrence. This study considered six types of emotional states (joy, anger, sadness, surprise, fear, and disgust) exported from the Face Reader software. They were collected with a frame rate of 29.97; a deviation of  $\pm 2000$  ms was used for calculating the proportion of recurrence. This time frame was chosen based on a preliminary investigation, which showed that this lag was optimal for the task. The recurrence of the types of emotions determined the degree to which both learners were in the same emotional state.

The facial expressions were not labeled when the learners were talking and were removed when either of them was talking, considering that the baseline of the calibrations for the software was trained when the mouths were closed. Using data from the participants would render the calibration inaccurate. In addition, two coders checked

the reliability of the automated coding by randomly selecting, manually coding, and checking the accuracy of the automatically detected emotional states. The accuracy of the recognition system was 78%. These values (emotional states) calculated from the emotional state was used for prediction analysis of the learning process and the performance.

## Results

### Effects of Using PCA Facilitation

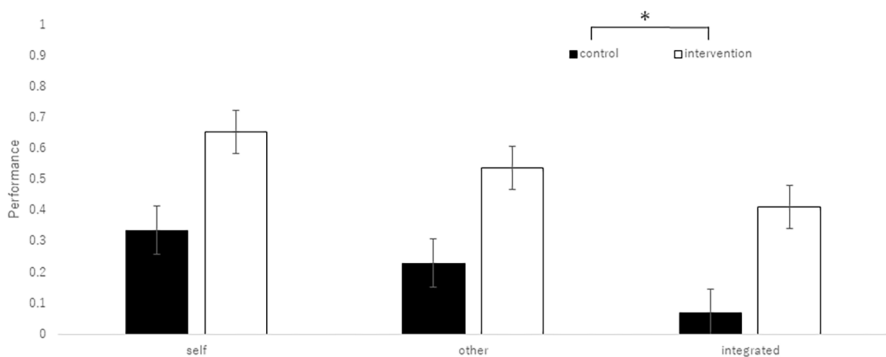
#### Collaborative Performance Under Intervention and Control Conditions

For the analysis of learning performance, a  $2 \times 3$  mixed-way analysis of variance (ANOVA) was conducted. Figure 6 shows the average gain score results for each type of gain score used as a dependent variable. No interaction was noted between the two factors ( $F(2,84)=0.091$ ,  $p=0.913$ ,  $\eta_p^2 = 0.002$ ). The main effects of the experimental conditions revealed that the score when using the intervention condition was higher than that when using the control condition ( $F(1,42) = 32.165$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.511$ ).

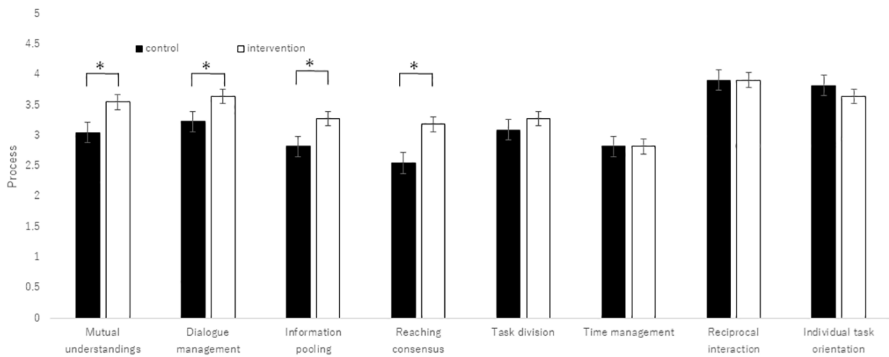
The results indicated that regardless of the test type, higher performance was achieved under the intervention condition than under the control condition. Thus, the result supported hypothesis H1-1.

#### Collaborative Process Under Intervention and Control Conditions

Similarly, a  $2 \times 8$  mixed-way ANOVA was conducted for the analysis of the collaborative process. Figure 7 shows the average gain score results for each type of gain score used as a dependent variable. An interaction between the two factors was noted ( $F(7,294)=4.051$ ,  $p=0.001$ ,  $\eta_p^2 = 0.087$ ). Further analysis on the simple main effects revealed that the scores for the intervention condition were higher than those for the



**Fig. 6** Average gain score for each condition. The error bar indicates the standard error, and \* indicates statistical significance



**Fig. 7** Average score of the collaborative process under each condition. The error bar indicates the standard error, and \* indicates statistical significance

**Table 1** Results of regression analysis on collaborative performance under control condition

Type	$R^2$	F-value	p-level significance
Self	−0.098	0.370	0.775
Other	0.231	1.810	0.181
Integrated	0.075	0.488	0.694

control condition in the categories of “mutual understanding,” “dialog management,” “information pooling,” and “reaching consensus” ( $F(1,336) = 6.976, p < 0.001, \eta_p^2 = 0.874$ ;  $F(1,336) = 4.670, p < 0.001, \eta_p^2 = 0.823$ ;  $F(1,336) = 5.765, p < 0.001, \eta_p^2 = 0.852$ ;  $F(1,336) = 11.300, p < 0.001, \eta_p^2 = 0.918$ ).

The results showed that interventions from the PCA facilitated the learning processes investigated in this study. The results were consistent with those of previous studies (Hayashi, 2018, 2019a), showing that facilitation by the PCA enabled the collaborative process, which supported H1-2. Moreover, the learning process facilitated by the PCA such as “mutual understanding,” “dialog management,” “information pooling,” and “reaching consensus” were interpreted as important processes for facilitating learning gain, as shown in Fig. 6.

## Using Indicator Recurrence for Predictions: Gaze, Lexical Conformance, and Facial Gestures

### Predicting Performance Using Recurrence of the Three Indicators

To investigate H2-1 and determine whether the recurrence of indicators facilitates collaborative performance, multiple regression analysis was conducted. In this analysis, the three variables of recurrence (gaze, lexical, and facial), which were determined to be important variables that influence performance, were used. Multiple regression was conducted for each of the three types of performance, self-concept, other-concept, and integrated. Tables 1 and 2 present the results obtained from the regression analysis under the control and intervention conditions, respectively.

**Table 2** Results of regression analysis on collaborative performance under intervention condition

<i>Type</i>	<i>R</i> <sup>2</sup>	<i>F</i> -value	<i>p</i> -level significance
Self	0.159	1.136	0.360
Other	0.144	1.012	0.410
Integrated	0.234	1.841	0.175

The regression analysis results did not indicate any relation between the degree of recurrence and performance. In the next section, these indicators are further investigated as a part of the collaborative process.

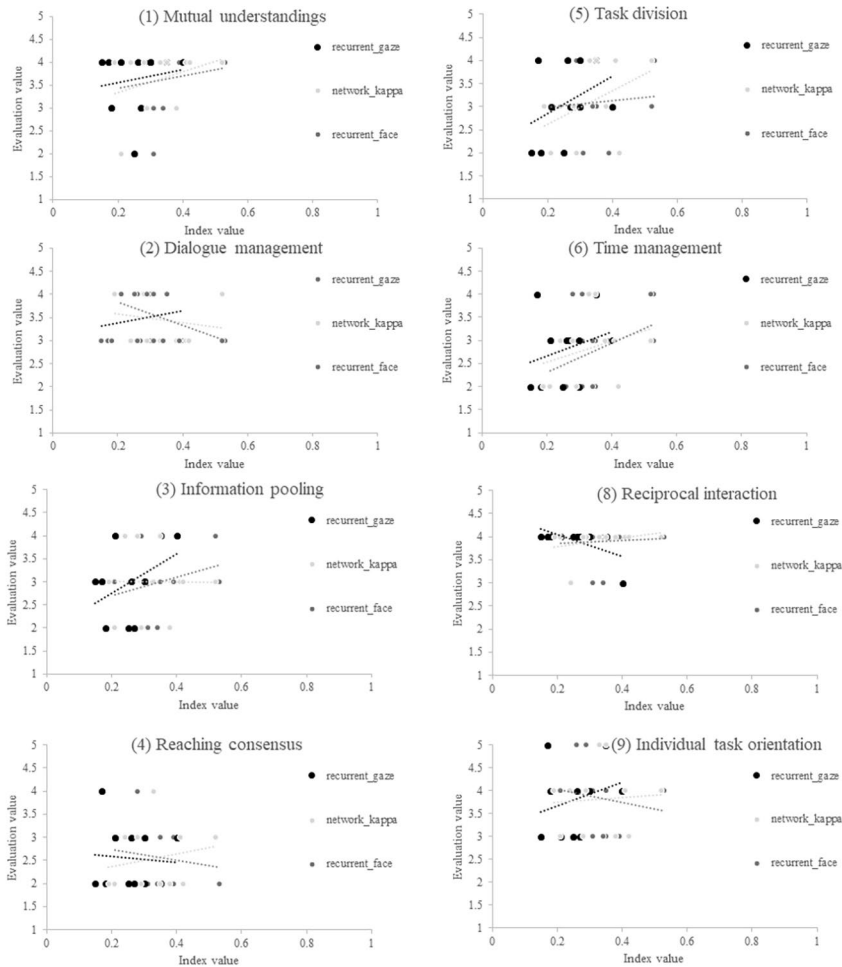
### Predicting Process Using the Recurrence of the Three Indicators

To investigate H2-2, that is, to determine whether the recurrence of indicators facilitates the collaborative process, multiple regression analysis was conducted. Again, in this analysis, all three variables of recurrence (gaze, lexical, and facial) were used. Multiple regression analysis was conducted for each of the eight types of collaborative processes, namely, “mutual understanding,” “dialog management,” “information pooling,” “reaching consensus,” “task division,” “time management,” “reciprocal interaction,” and “individual task orientation.” Figs. 8 and 9 show the correlations between the main results under the control and intervention conditions, respectively. Tables 3 and 4 present the statistical results obtained from the regression analysis under the control and intervention conditions, respectively. The statistical analysis results showed the significance of regression.

The overall regression analysis (including both conditions) showed that the recurrence predictor variables were able to detect the collaborative process. When focusing on different conditions, it was found that the number of significant predictions differed between the control and intervention conditions. Only three out of the eight collaborative processes were predicted under the control condition, whereas six out of the eight processes were predicted under the intervention condition. These results supported hypothesis H2-2 and indicated that the recurrence predictors were more useful under conditions where learning was facilitated by the PCA. As predicted, the facilitation provided to learners with suggestions on ways to coordinate with each other, enabling them to synchronize their behavior, which resulted in high correlation during the learning process.

### Qualitative Analysis of the Collaborative Process

In this study, learners interacted with each other, while the PCA provided prompts to facilitate the learning process in a knowledge integration explanation task. In an ideal process, learners would (1) take turns and explain their concepts by reading the text, (2) further ask questions until they both reach a common understanding, and (3) further develop abstract understanding and knowledge that can be used to explain the phenomenon. The results from the previous sections show that the indexes were able to detect the learning process when using the PCA. This section further qualitatively describes the differences in the collaboration process between the two conditions. An



**Fig. 8** Correlation between the collaborative process and the three types of recurrence in the control condition

example of a transcript of the initial phase, which is critical for developing mutual understanding, is shown below.

<Intervention condition>.

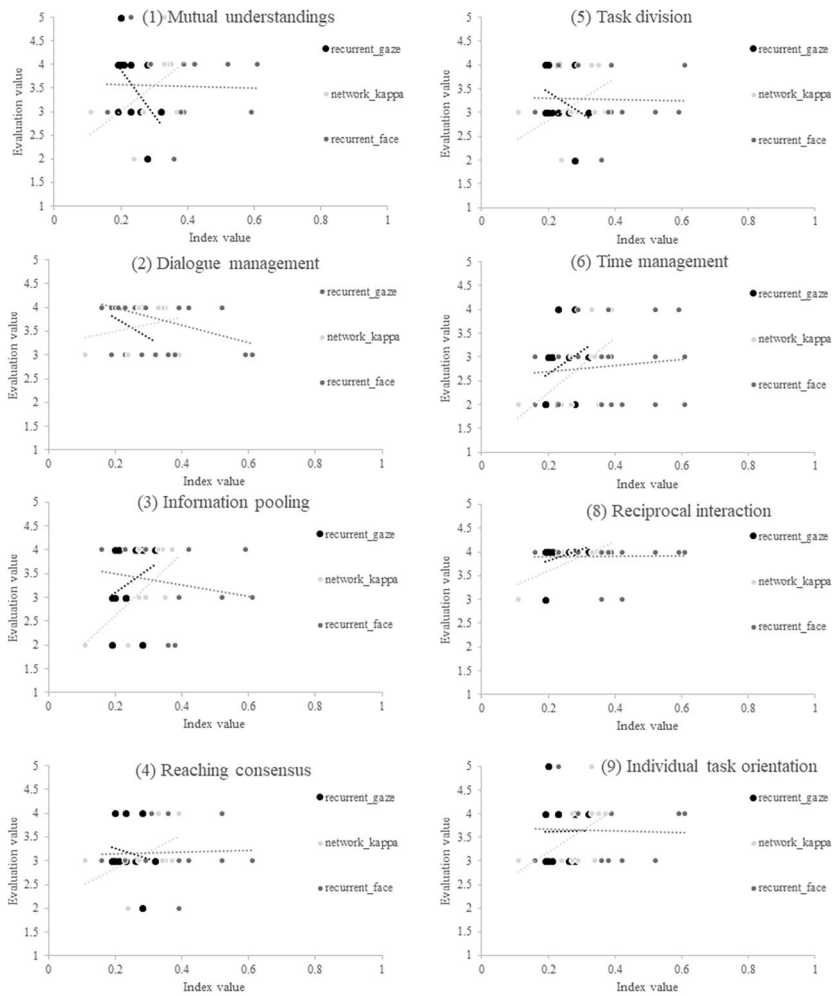
PCA: As a reminder, the goal of the task is to provide explanations about each other's knowledge and integrate those. [facilitating the purpose]

A: So, from your explanation I understood that top-down processing is something that is related to individual's past experience. Is that correct?

B: As far as what I read and see on the text written on the screen here, yes, using knowledge acquired from experience is the first step <gaze moving to the same area>.

A: I see. So, using knowledge from experience is important as it will trigger how we see things.





**Fig. 9** Correlation between the collaborative process and the three types of recurrence in the intervention condition

B: Yea, that is called using a schema. The knowledge that I was talking about when I read the description.

PCA: Good job, you two. Schema is a one of the key terms. Try to further think how it is useful for explaining the integrated explanation. [facilitating based on the detection of a keyword; metacognitive facilitation]

B: So, it seems that we are on track and perhaps let us share each other's ideas, so we can make a whole integrated explanation?

A: Sure, so let me further explain about what is written on the screen on my side. [reciprocal]

<Control condition>.

**Table 3** Results of regression analysis on collaborative process, control condition

Type	$R^2$	F-value	p-level significance
Mutual understanding	0.163	1.177	0.346
Dialog management	0.298	2.547	0.088
Information pooling	0.359	3.368	<0.05*
Reaching consensus	0.192	1.430	0.266
Task division	0.479	5.517	<0.05*
Time management	0.242	1.916	0.163
Reciprocal interaction	0.632	10.329	<0.001**
Individual task orientation	0.233	1.830	0.177

\*indicates significance at 0.05,

\*\*indicates significance at 0.01

**Table 4** Results of regression analysis on collaborative process, intervention condition

Type	$R^2$	F-value	p-level significance
Mutual understanding	0.729	16.220	<0.001**
Dialog management	0.602	9.089	<0.001**
Information pooling	0.636	10.510	<0.001**
Reaching consensus	0.306	2.649	0.080
Task division	0.503	6.078	<0.05*
Time management	0.378	3.652	<0.05*
Reciprocal interaction	0.724	15.817	<0.001**
Individual task orientation	0.342	3.122	0.051

\* indicates significance at 0.05,

\*\*indicates significance at 0.01

A: Do you have any idea about how to interpret the concept? Or should I just read the text again?

B: I think that it's okay to think on our own and simply talk about what we felt about the concept. Let us just read separately.

A: Okay. [reading....]

B: As far as what I'm reading now, the process of bottom-up may be related to something to collect data and then make decisions based on them.

A: So, these two concepts are different?

B: I guess so. Maybe we have already reached the conclusion and achieved the goal. The two are just different.

From these sample dialogs, it can be observed that under the intervention condition, the learners first received facilitation about being aware of the purpose of the study. Then, they followed the instructions and started to converse about how to establish common ground for the two different concepts. During this process, both learners' gazes were directed on the same area, they were using similar wordings (in this case "experience"), and they both smiled when they agreed during their exchange of thoughts. However, under the control condition, learners worked relatively individually without a grounding process as that observed under the intervention condition. Moreover, when the PCA provided metacognitive suggestions, the learners further took turns to achieve their goals on integrating their perspectives. No cooperative activity was observed under the control condition.

These two representative examples showed that under the intervention condition, the learners had higher occurrence of synchronized gaze, language use, and facial expressions, which were associated with the collaborative process of developing

mutual understanding and reciprocal interaction. These qualitative data supported the quantitative data that were shown in the previous sections, related to the correlations between the collaborative process and the three types of recurrences under the intervention condition.

## Discussion

### Interpretation of the Results

The analytical results confirmed that the manipulation check in the experiment was appropriate. In particular, the intervention condition that used facilitation by the PCA more effectively promoted learning performance than the control condition without the PCA. Unlike the study of Hayashi (2019a), the present study conducted direct comparisons between the two conditions and provided further evidence on how the use of the PCA influenced learning performance, processes, and synchronization detection. This indicates that synchronization metrics can be useful in combination with the PCA intervention; further verification of this result will be conducted in future studies. Compared to Hayashi (2019a), the learners' explanatory activities were divided into three categories, including self-concept, other-concept, and integration. Data analysis was performed accordingly, and it was shown that the PCA influenced all three types of perspective explanations. This shows that the PCA had a facilitative effect on the learners' explanatory activities. Upon examination, some results were consistent with those of Hayashi (2019a) using the same dependent variables for the collaborative learning process. Specifically, in Hayashi (2019a), the effect of using the PCA was observed in "reaching consensus" and "task division;" however, in the present study, there were more areas with significant impact, including "sustaining mutual understanding," "dialog management," "information pooling," and "reaching consensus."

The results of this study showed that it was difficult to estimate the performance of the task content using recurrence because no significant difference was observed for any of the items. This can be interpreted in several ways. The recurrence indexes and the evaluation of the process are both variables that focus on the interaction and process at a group level. However, learning performance is an index evaluating the understanding at the individual level. This gap could have caused the difficulty in detecting performance using the recurrence indexes. Another possibility is that even some learners who interacted with good recurrence and achieved good co-construction could have simply failed on the post-test. Because the performance evaluation was conducted after the main discussion task, there was a time lag between the point of evaluation of the process and post-test. After the discussions in the main task, learners might have reconsidered their knowledge and modified it using their naivety; thus, the process might not have reflected the performance.

Furthermore, the number of processes that can be predicted was higher with PCA than without PCA. This indicates that state detection using the three recurrent indicators is feasible under the condition that the collaborative process is likely to occur due to the facilitation of the PCA. These findings suggest that including various types

of indexes that are discovered in human–human interaction studies are also useful for detecting the collaborative learning process that requires coordination in tasks, such as that utilized in this study. Further studies for the development and design of tutoring systems to detect the type of learners' interaction process may incorporate indexes for synchronizations and further investigate the different types of verbal and nonverbal indexes. Moreover, the future development of PCAs will involve the use of a combination of these indexes with PCAs that provide metacognitive suggestions. The automated detection of synchronizations can be used for providing real-time feedback on the state of synchronization of the dyad to inform them about their level of coordination.

Finally, a debate arises on whether the PCA has a direct influence on the three indicators and can be biased toward the information given instead of their own collaborative processes. This debate is attributed to the fact that the intervention condition can have more data as additional time is spent in reviewing the PCA feedback. Although some learners can have such reactions toward PCA interventions, the data showed that the influence was limited. Checks were made to verify the presence of any difference between the two conditions in terms of the number of (1) gaze plots, (2) utterances, and (3) facial expressions. For (1), the variance analysis of the between-subject factor for one factor revealed no differences between the conditions ( $F(1, 43)=0.217$ ,  $p=0.643$ , and  $\eta_p^2=0.005$ ). For (2), the variance analysis of the one between-subject factor showed no difference between the conditions ( $F(1, 43)=0.106$ ,  $p=0.746$ , and  $\eta_p^2=0.002$ ). For (3), the variance analysis of the between-subject factor for one factor revealed no differences between the conditions ( $F(1, 43)=0.586$ ,  $p=0.44$ , and  $\eta_p^2=0.014$ ). Thus, the post-hoc analysis confirmed that the direct reaction of participants did not influence the number of data that were analyzed for comparison.

## Predicting Learning Process and Contributions of this Study

Research on learning support using computers has been conducted for several years (Biswas et al., 2005; Koedinger et al., 1997; Leelawong & Biswas, 2008); for example, Cognitive Tutor (Koedinger et al., 1997), which is known as a typical ITS in research, monitors each step in detail when a learner solves a problem and provides appropriate support according to the learner's situation (Koedinger et al., 2013). This research provides valuable insights into the design of such adaptive learning support systems. Particular, this study proposed an index to detect the quality of the learners' process based on data collected from sensors and cameras, which were mounted on the computer, and from audio through microphone. Using these indicators, the process of learner interaction was monitored in real time to determine if the learners were on the path, such as successfully coordinating during their explanatory activity. For example, if the level of synchronization during interaction did not meet the expectations, the PCA provided suggestions to the learners to focus on the same part of the task and consider that part carefully by providing hints of common words that could be used for explanations. However, automatic classification of the learners' situation in real time using these three indicators remained a challenge for future work. Particularly, the challenge is to design an algorithm to classify the learners' situation using a machine learning method. Therefore, the feasibility of such an approach can

be further examined using the findings obtained in this study. Discriminant analysis that can be classified through supervised learning was performed with a focus on the six types of processes that were estimated by the regression analysis, under the condition of using the PCA (i.e., “mutual understanding,” “dialog management,” “information pooling,” “task division,” “time management,” and “reciprocal interaction”). The results of the analysis are summarized in Table 5.

Relatively high values were obtained for dialog management (81.8%), information pooling (90.9%), and reciprocal interaction (99%), demonstrating the usefulness of classification using discriminant analysis. Further, researchers can take advantage of these indexes to discover new patterns in existing datasets in collaborative learning tasks, such as jigsaw learning tasks, which require coordination. Considering that many learning tasks require aspects of coordination, especially in learning environments with low awareness, the index for synchronizations can provide a wider frame of whether learners are interacting effectively. In particular, the three indexes found in these classifications suggest that verbal and nonverbal cues, which can be collected by easily accessible devices (camera, voice recordings, and cheap sensing devices), can be used to predict multiple types of collaboration processes. Further studies using machine learning may extend the notions of this study and further determine what types of parameters may best predict the learning process. Moreover, different indexes and focus on different stages of the collaborative activities that best fit the learners’ process can be tested. Additionally, this study provides fundamental knowledge on the types of real-time detection modules that can be implemented in tutoring systems with PCAs that aim to provide adaptive feedback based on the degree of success of a collaborative process. Past studies in tutoring systems that have focused on individual learning have developed numerous cognitive models; however, there are relatively few studies that focus on modeling learners’ states at the pair level where the success of interaction plays an important role in achieving the task. Thus, this study investigated the idea of the use of an index of multimodal synchronizations, which was studied in cognitive science, for modeling the process of learners at the pair level and provides new implications for future studies. Another contribution of this study is that this index can be used for evaluation of the quality of learning process, especially for determining the degree of the learners’ coordination during collaborative learning. Teachers may also use such detection methods to develop real-time monitoring systems for classrooms in online courses to monitor learners’ performances.

### Limitations of this Study

This study has several limitations that should be addressed. The PCA was intended to facilitate metacognitive processes during the task; however, concerns that such interventions can distract the learners from their free conversations and independent

**Table 5** Classification results by the discriminant analysis of major items using the PCA.

	Sustaining mutual understanding	Dialog management	Information pooling	Task division	Time management	Reciprocal interaction
PCA	63.6%	81.8%	90.9%	72.7%	63.6%	99%

activities exist. Therefore, all transcripts were checked to determine if the PCA comments drastically changed the learners' conversations (e.g., the learners' conversation were completely stopped or the learners waited for the PCA or listened only to the PCA); however, there were no such instances. Although such intervention seems not to have significantly altered the flow of conversation, it may have served as a distraction and affected the natural semantic richness of the conversational expressions used to some extent. Developing a system to complement the conversations is a challenging research topic, and future studies should further examine how to facilitate independent interactions of learners.

Regarding the recurrence index of information overlap, this study only focused on the use of the same words. It is possible to perform matching at the semantic level; nevertheless, it was out of the scope of this study, which aimed at detecting the recurrences between the contents of the conversations. Natural language processing and word vector space mapping are potential ways to capture the type of words that were used during the conversations.

As mentioned in Sect. 1.2.3, previous studies have identified the limitations of using facial expressions to detect emotional states. Facial expressions, which were used in this study, do not cover all the methods for capturing the emotional states of synchronization. Therefore, the use of only on facial expressions (action units) for detecting emotional states should be verified by using other methods. Additionally, future studies may further examine other synchronized embodied gestures, such as body gestures, for detecting emotional states using motion capturing or other methods. Finally, there are ethical ramifications of using facial recognition for emotion detection. Developers of models using facial recognition technology should consider such issues. Moreover, when analyzing facial expressions during a conversation, it is not desirable for the shape of the corners of the mouth to change in an accidental manner, and the mouth movements can be considered to include some noise. The present study conducted a post-hoc analysis to ensure that this noise was minimal, and the data that were collected during the conversations were mostly undetectable or labeled as neutral. However, it can be argued that the analysis can be performed even while learners are talking, if the focus is on a macro-level of facial expressions, such as considering a certain time span. Although noise would be added, the proposed approach could allow meaningful analysis if compared with other situations that involve the same speakers. It is problematic to conduct a direct comparison between a non-speaker and speaker; however, a plausible approach is to consider the macro-level analysis of speakers. Moreover, this study assumed the creation of a learning support system that systematically calculates recurrent data automatically detected by sensors. Therefore, the manual modification of the obtained data should be avoided as much as possible. Future studies should consider investigating the facial expressions of learners while talking as an index.

Another consideration is the generalization of the findings toward other types of collaborative activities. The task used in this study includes the fundamental process of knowledge integration and explanation activities that are used in jigsaw learning. This task was used to enable more effective capture of the process of joint actions and referential attention of different perspectives. Although this may be considered task dependent, the author considers that the findings may be generalized to other

types of collaborative activities that require the learning of tasks through coordination and mutual understanding. Collaborative learning activities requiring this type of interaction can be found in various types of active learning and online-based learning environments.

This study was conducted in a laboratory-based controlled experiment, and the learning task that was used was simpler and less realistic than that used in a traditional learning setting in learning sciences. The validity of results collected from experimental data is often questioned in laboratory experiments that use simplified learning tasks. The advantages of using a simple experiment include the ability to determine what is happening under certain circumstances and to eliminate the noise of different factors, which may be masking the interactions. Moreover, this study selected a non-face-to-face situation because it aimed to investigate how learners develop common understanding and coordinate with each other in environments with fewer social cues, such as in online learning. The findings of this study provide basic and scientific results on the proposed synchronization indexes, which can help to successfully detect situations in which coordination is required. To validate this, future studies should use naturalistic tasks and determine how the recurrences can be applied in environments in which coordination is required, such as in jigsaw-based learning activities in classrooms.

Moreover, there are several important interaction processes that this study was not able to focus on. These included applied processes, such as knowledge co-construction (Beers et al., 2005), social interactions (Weinberger & Fischer, 2006), and regulatory interactions (Hadwin et al., 2018). These processes were not covered because this study focused on the relationship between synchronized coordination and the collaborative process during communication in computer-mediated tasks. Therefore, interactions were limited to learners interacting under low-awareness circumstances. Thus, future research directions should include further examinations in applied settings and face-to-face environments, which may include third variables and noise.

Finally, considering that the number of participants in this study was small (11 dyads), some of the non-significant comparisons, such as those presented in Fig. 7, may differ when analyzed using a larger amount of data. Moreover, it is a challenge to investigate the effect of the diversity of the learners (e.g., age and educational level) on the results obtained from the analysis conducted in this study. Therefore, collecting more data with higher diversity is another challenge for the future.

## Conclusions

Providing explanations to others and the development of common ground to make sense of contents are beneficial for deepening knowledge and for generating abstract understandings. Despite the ideal process for such interactions, successful coordination with peers may be challenging for novice learners, and these activities become more difficult in environments with low awareness, such as in distant learning environments and online communication tasks. Several studies have been reported on tutoring systems that detect learners' interactions, and the author has also conducted several studies utilizing PCAs. However, there is a lack of investigation on the devel-



opment of indexes that measure the extent to which learners coordinate and establish common ground in the initial process of knowledge integration explanation tasks. Therefore, determining the types of indexes that can be used to capture the successful mutual coordination of learners and understanding those indexes is important for predicting the learning process. Past studies in psychology have identified the degree of pair level of coordination, but few studies have developed indexes for designing tutoring systems for collaborative learning. Therefore, this study considered three types of indicators—gaze, linguistic expression, and emotion—to understand recurrence and examined whether each would be effective in detecting collaborative learning performance and processes. Gaze recurrence, which measures the degree of coincidence of learners' gaze, and the number of language alignments that used the same expression were calculated. The results showed that the use of a PCA promoted learning performance and collaborative processes compared with instances without a PCA. Subsequently, under the condition where a PCA was used, many predictions could be made for the dependent variables of the collaborative learning process using the three proposed indicators. This finding shows that the suggested indexes of recurrence are useful indicators for detecting the process of collaborative learning.

Moreover, the findings indicate that suggestions from PCAs lead to better prediction of learning process in areas such as communication, joint information processing, coordination, and motivation. These results contribute to the knowledge that combinations of the multimodal indexes of synchronization collected by noncontact devices are useful for predicting learning process in collaborative tasks that require coordination. The findings also open the door for possible studies on synchronization in the field of social psychology and social signal processing, which may be further incorporated for studying, developing, and designing ITSs for collaborative support. Future studies may focus on real-time detection using PCAs that capture learners that are on the path of adequate coordination process.

**Acknowledgements** This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid for Scientific Research (B), Grant Number JP20H04299 and JP23H03510.

**Data Availability** All data generated or analyzed during this study are included in this published article.

**Code Availability** Not applicable.

## Declarations

**Conflict of Interest** The author declares no conflicts of interest.

**Ethics Approval** The study was conducted after institutional ethical review and approval by the ethical reviewing committee of Ritsumeikan University (Kinugasa-2020-7).

**Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alkilani, H. A., & Nusir, I. M (2022). Students' performance-prediction-Model based on Physical and Physiological Constraints. 2022 *19th International Multi-Conference on Systems, Signals & Devices (SSD)*, Sétif, Algeria, 2022, 47–52. <https://doi.org/10.1109/SSD54932.2022.9955809>
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207. [https://doi.org/10.1207/s15327809jls0402\\_2](https://doi.org/10.1207/s15327809jls0402_2)
- Apperly, I. (2018). Mindreading and psycholinguistic approaches to perspective taking: Establishing common ground. *Topics in Cognitive Science*, 10(1), 133–139. <https://doi.org/10.1111/tops.12308>
- Aronson, E., & Patnoe, S. (1997). *The Jigsaw Classroom: Building Cooperation in the Classroom* (2nd ed.). Addison Wesley Longman.
- Asterhan, S. C., & Schwarz, B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist*, 51(2), 164–187. <https://doi.org/10.1080/00461520.2016.1155458>
- Azevedo, R. (2020). Reflections on the field of metacognition: Issues, challenges, and opportunities. *Meta-cognition and Learning*, 15, 91–98. <https://doi.org/10.1007/s11409-020-09231-x>
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96(3), 523–535. <https://doi.org/10.1037/0022-0663.96.3.523>
- Azevedo, R., Behnagh, R. F., Duffy, M., Harley, J. M., & Trevors, G. (2013). Metacognition and self-regulated learning in student-centered learning environments. In S. Land, & D. Jonassen (Eds.), *Theoretical foundations of student-center learning environments* (2nd ed.). Routledge.
- Baltrušaitis, T., Robinson, P., & Morency, L. (2016). OpenFace: An open source facial behavior analysis toolkit. 2016 *IEEE Winter Conference on Applications of Computer Vision*, 1–10. <https://doi.org/10.1109/WACV.2016.7477553>
- Bartlett, S. M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing facial expression: machine learning and application to spontaneous behavior. in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2, 568–573, IEEE, San Diego, CA, USA, June 2005.
- Beers, P. J., Boshuizen, H., Kirschner, P. A., & Gijssels, W. H. (2005). Computer support for knowledge construction in collaborative learning environments. *Computers in Human Behavior*, 21(4), 623–643. <https://doi.org/10.1016/j.chb.2004.10.036>
- Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new paradigm for educational software. *Applied Artificial Intelligence*, 19(3–4), 363–392. <https://doi.org/10.1080/08839510590910200>
- Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121(1), 41–57. <https://doi.org/10.1016/j.cognition.2011.05.011>
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 22(6), 1482–1493. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Cai, Y., Shimojo, & Hayashi, S. (2020). Y. Observing Facial Muscles to Estimate the Learning State During Collaborative Learning: A Focus on the ICAP Framework. *Proceedings of the 28th International Conference on Computers in Education(ICCE2020)*, 119–126.
- Chi, M. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>

- Chi, M., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Chi, M., Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7)
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). APA Press. <https://doi.org/10.1037/10096-006>
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
- Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multi-modal learning analytics. *British Journal of Educational Technology*, 51(5), 1441–1449. <https://doi.org/10.1111/bjet.13015>
- D'Angelo, S., & Schneider, B. (2021). Shared gaze visualizations in collaborative interactions: Past, present and future. *Interacting with Computers*, 33(2), 115–133. <https://doi.org/10.1093/iwcomp/iwab015>
- D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170. <https://doi.org/10.1016/j.learninstruc.2012.05.003>
- Dale, R. (2015). An integrative research strategy for exploring synergies in natural language performance. *Ecological Psychology*, 27(3), 190–201. <https://doi.org/10.1080/10407413.2015.1068649>
- Dignath, C., & Buttner, G. (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and secondary school mathematics classes—insights from video-based classroom observations and teacher interviews. *Metacognition and Learning*, 13, 127–157. <https://doi.org/10.1007/s11409-018-9181-x>
- Engelmann, B. J., & Pogossyan, M. M. (2013). Emotion perception across cultures: The role of cognitive mechanisms. *Frontiers in Psychology*, 118(4), <https://doi.org/10.3389/fpsyg.2013.00118>
- Galati, A., & Brennan, E. S. (2021). What is retained about common ground? Distinct effects of linguistic and visual co-presence. *Cognition*, 215, 10489. <https://doi.org/10.1016/j.cognition.2021.104809>
- Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, 45(4), 234–244. <https://doi.org/10.1080/00461520.2010.515933>
- Graesser, A., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618. <https://doi.org/10.1109/TE.2005.856149>
- Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In P. A. Alexander, D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 83–106). Routledge. <https://doi.org/10.4324/9781315697048-6>
- Hayashi, Y. (2012). On pedagogical effects of learner-support agents in collaborative interaction. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems (ITS 2012) Lecture Notes in Computer Science, vol 7315* (pp. 22–32). Springer. [https://doi.org/10.1007/978-3-642-30950-2\\_3](https://doi.org/10.1007/978-3-642-30950-2_3)
- Hayashi, Y. (2018). The power of a maverick in collaborative problem solving: An experimental investigation of individual perspective-taking within a group. *Cognitive Science*, 42(S1), 69–104. <https://doi.org/10.1111/cogs.12587>
- Hayashi, Y. (2019a). Towards supporting collaborative learning with an intelligent tutoring system: Predicting learning process by using gaze and verbal information. *Cognitive Studies*, 26(3), 343–356. <https://doi.org/10.11225/jcss.26.343> (In Japanese).
- Hayashi, Y. (2019b). Multiple pedagogical conversational agents to support learner-learner collaborative learning: Effects of splitting suggestion types. *Cognitive Systems Research*, 54, 246–257. <https://doi.org/10.1016/j.cogsys.2018.04.005>
- Hayashi, Y. (2019c). Detecting collaborative learning through emotions: An investigation using facial expression recognition. In A. Coy, Y. Hayashi, & M. Chang (Eds.), *Intelligent Tutoring Systems. ITS 2019. Lecture Notes in Computer Science, vol 11528* (pp. 89–98). Springer. [https://doi.org/10.1007/978-3-030-22244-4\\_12](https://doi.org/10.1007/978-3-030-22244-4_12)
- Hayashi, Y. (2020). Gaze awareness and metacognitive suggestions by a pedagogical conversational agent: An experimental investigation on interventions to support collaborative learning process and performance. *International Journal of Computer-Supported Collaborative Learning*, 15, 469–498. <https://doi.org/10.1007/s11412-020-09333-3>

- Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1), 27–54. <https://doi.org/10.1016/j.edurev.2010.07.004>
- Järvelä, S., & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. *Teachers College Record*, 113(2), 350–374. <https://doi.org/10.1177/016146811111300205>
- Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39(10), 1123–1134. <https://doi.org/10.1037/0003-066X.39.10.1123>
- Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, 9(2), 187–215. <https://doi.org/10.1007/s11409-014-9112-4>
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30–43.
- Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, 342(6161), 935–937. <https://doi.org/10.1126/science.1238056>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's brain system. *International Journal of Artificial Intelligence in Education*, 18(3), 181–208.
- Martinez-Maldonado, R., Kay, J., Yacef, K., & Schwendimann, B. (2012). An interactive teacher's dashboard for monitoring multiple groups in a multi-tabletop learning environment. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems. ITS 2012. Lecture Notes in Computer Science*, vol 7315 (pp. 482–492). Springer. [https://doi.org/10.1007/978-3-642-30950-2\\_62](https://doi.org/10.1007/978-3-642-30950-2_62)
- Matsuda, N., Cohen, W. W., & Koedinger, K. R. (2015). Teaching the teacher: Tutoring SimStudent leads to more effective cognitive tutor authoring. *International Journal of Artificial Intelligence in Education*, 25, 1–34. <https://doi.org/10.1007/s40593-014-0020-1>
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 63–86. <https://doi.org/10.1007/s11412-006-9005-x>
- Menenti, L., Pickering, M. J., & Garrod, S. C. (2012). Toward a neural basis of interactive alignment in conversation. *Frontiers in Human Neuroscience*, 6, 185. <https://doi.org/10.3389/fnhum.2012.00185>
- Miyake, N., & Kirschner, P. A. (2014). The social and interactive dimensions of collaborative learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 418–438). Cambridge University Press.
- Moridis, N. C., & Economides, A. A. (2012). Affective learning: Empathetic agents with emotional facial and tone of voice expressions. *IEEE Transactions on Affective Computing*, 3(3), 260–272. <https://doi.org/10.1109/T-AFFC.2012.6>
- Nalls, A. J., & Wickerd, G. (2022). The jigsaw method: Reviving a powerful positive intervention. *Journal of Applied School Psychology*. <https://doi.org/10.1080/15377903.2022.2124570>
- Ochoa, X., & Worsley, M. (2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, 3(2), 213–219. <https://doi.org/10.18608/jla.2016.32.10>
- Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347. <https://doi.org/10.1017/S0140525X12001495>
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92(3), 544–555. <https://doi.org/10.1037/0022-0663.92.3.544>
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, 18(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2007.01914.x>
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007b). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891. <https://doi.org/10.1016/j.humov.2007.07.002>
- Richardson, C. J., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior*, 71, 402–417. <https://doi.org/10.1016/j.chb.2017.02.001>
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95(2), 240–257. <https://doi.org/10.1037/0022-0663.95.2.240>

- Roscoe, D. R., Segedy, R. J., Sulcer, B., Jeong, H., & Biswas, G. (2013). Shallow strategy development in a teachable agent environment designed to support self-regulated learning. *Computers & Education*, 62, 286–297. <https://doi.org/10.1016/j.compedu.2012.11.008>
- Rosé, C. P., & Ferschke, O. (2016). Technology support for discussion based learning: From computer supported collaborative learning to the future of massive open online courses. *International Journal of Artificial Intelligence in Education*, 26, 660–678. <https://doi.org/10.1007/s40593-016-0107-y>
- Rosé, C., Wang, Y., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237–271. <https://doi.org/10.1007/s11412-007-9034-0>
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4(1), 69–92. <https://doi.org/10.1007/s11412-008-9054-4>
- Ryskin, R., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L., & Nguyen, E. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language*, 74, 46–76. <https://doi.org/10.1016/j.jml.2014.04.003>
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3(3), 265–283. [https://doi.org/10.1207/s15327809jls0303\\_3](https://doi.org/10.1207/s15327809jls0303_3)
- Schneider, B., & Pea, R. (2014). Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning*, 9, 371–395. <https://doi.org/10.1007/s11412-014-9202-y>
- Schneider, J., Börner, D., Van Rosmalen, P., & Speccht, M. (2015). Augmenting the senses: A review on sensor-based learning support. *Sensors (Basel, Switzerland)*, 15(2), 4097–4133. <https://doi.org/10.3390/s150204097>
- Schneider, B., Hassan, J., & Sung, G. (2022). Augmenting social science research with multimodal data collection: The EZ-MMLA Toolkit. *Sensors (Basel, Switzerland)*, 22(2), 568. <https://doi.org/10.3390/s22020568>
- Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, 47(1), 1–24. [https://doi.org/10.1016/0010-0277\(93\)90060-9](https://doi.org/10.1016/0010-0277(93)90060-9)
- Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Eds.), *Spatial language and dialogue* (pp. 23–39). Oxford University Press.
- Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, 2(1), 13–48. <https://doi.org/10.18608/jla.2015.21.3>
- Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 326–332. <https://doi.org/10.1037/0096-1523.29.2.326>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Skiendziel, T., Rösch, G. A., & Schultheiss, C. O. (2019). Assessing the convergent validity between the automated emotion recognition software noldus facereader 7 and facial action coding system scoring. *PLoS One*, 14(10), <https://doi.org/10.1371/journal.pone.0223905>. Article ID e0223905.
- Soliman, D., Costa, S., & Scardamalia, M. (2021). Knowledge building in online mode: Insights and reflections. *Education Sciences*, 11(8), 425. <https://doi.org/10.3390/educsci11080425>
- Stewart, A., Keirn, Z., & D'Mello, S. K. (2021). Multimodal modeling of collaborative problem-solving facets in triads. *User Modeling and User-Adapted Interaction*, 31(4), 713–751. <https://doi.org/10.1007/s11257-021-09290-y>
- Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 50, 1446–1460. <https://doi.org/10.3758/s13428-017-0996-1>
- Towne, W. B., Rosé, C. P., & Herbsleb, J. D. (2017). Measuring similarity similarly: LDA and human perception. *ACM Transactions on Intelligent Systems and Technology*, 8(1), 7. <https://doi.org/10.1145/2890510>
- van de Sande, C. C., & Greeno, J. G. (2012). Achieving alignment of perspectival framings in problem-solving discourse. *Journal of the Learning Sciences*, 21(1), 1–44. <https://doi.org/10.1080/10580406.2011.639000>

- van Kuilenburg, V. H., Wiering, M., & den Uyl, M. (2005). A model-based method for automatic facial expression recognition. In J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge, & L. Torgo (Eds.), *Machine Learning: ECML 2005 Lecture Notes in Computer Science, vol 3720*, (pp. 194–205). Springer. [https://doi.org/10.1007/11564096\\_22](https://doi.org/10.1007/11564096_22)
- VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31(1), 3–62. <https://doi.org/10.1080/03640210709336984>
- Vygotsky, L. S. (1980). *The development of higher psychological processes*. Harvard University Press.
- Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence Education*, 24, 33–61. <https://doi.org/10.1007/s40593-013-0001-9>
- Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin Review*, 20(4), 766–772. <https://doi.org/10.3758/s13423-013-0396-1>
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95. <https://doi.org/10.1016/j.compedu.2005.04.003>
- Zeng, Z., Pantic, M., Roisman, G., I., & Huang, T., S (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39–58. <https://doi.org/10.1109/TPAMI.2008.52>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Yugo Hayashi<sup>1</sup>

✉ Yugo Hayashi  
yhayashi@fc.ritsumei.ac.jp

<sup>1</sup> College of Comprehensive Psychology, Ritsumeikan University, Ibaraki 567-8570, Osaka, Japan