Received: 30 August 2023 | Revised: 21 February 2024 | Accepted: 1 May 2024

DOI: 10.1111/jcal.13007

ORIGINAL ARTICLE

Journal of Computer Assisted Learning WILEY

A learning analytics-based collaborative conversational agent to foster productive dialogue in inquiry learning

Adelson de Araujo | **Pantelis M. Papadopoulos** | **Susan McKenney** | **Ton de Jong**

Department of Learning, Data Analytics and Technology, Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, The Netherlands

Correspondence
Adelson de Araujo, Department of Learning, Data Analytics and Technology, Faculty of Behavioural, Management and Social Sciences, University of Twente, PO Box 217, Enschede 7500 AE, The Netherlands.
Email: a.dearaujo@utwente.nl

Abstract

Background: Sustaining productive student–student dialogue in online collaborative inquiry learning is challenging, and teacher support is limited when needed in multiple groups simultaneously. Collaborative conversational agents (CCAs) have been used in the past to support student dialogue. Yet, research is needed to reveal the characteristics and effectiveness of such agents.

Objectives: To investigate the extent to which our analytics-based Collaborative Learning Agent for Interactive Reasoning (Clair) can improve the productivity of student dialogue, we assessed both the levels at which students shared thoughts, listened to each other, deepened reasoning, and engaged with peer's reasoning, as well as their perceived productivity in terms of their learning community, accurate knowledge, and rigorous thinking.

Method: In two separate studies, 19 and 27 dyads of secondary school students from Brazil and the Netherlands, respectively, participated in digital inquiry-based science lessons. The dyads were assigned to two conditions: with Clair present (treatment) or absent (control) in the chat. Sequential pattern mining of chat logs and the student's responses to a questionnaire were used to evaluate Clair's impact.

Results: Analysis revealed that in both studies, Clair's presence resulted in dyads sharing their thoughts at a higher frequency compared to dyads that did not have Clair. Additionally, in the Netherlands' study, Clair's presence led to a higher frequency of students engaging with each other's reasoning. No differences were observed in students' perceived productivity.

Conclusion: This work deepens our understanding of how CCAs impact student dialogue and illustrates the importance of a multidimensional perspective in analysing the role of CCAs in guiding student dialogue.

KEYWORDS

collaborative learning, conversational agents, inquiry learning, learning analytics, productive dialogue

1 | INTRODUCTION

Collaboration is increasingly recognized as a vital 21st-century skill as it holds the potential for cognitive, social, and emotional gains for students of all levels (Hu & Chen, 2023; Noroozi et al., 2013). In this context, there is a growing emphasis on active learning methods, such as inquiry learning that can foster collaboration and the acquisition of deep conceptual knowledge (Chu et al., 2016). Inquiry learning encourages student participation, promotes familiarity with scientific practices, and cultivates high-level reasoning skills (Lazonder & Harmsen, 2016). In the inquiry process, dialogue can be essential for students to make common decisions and proceed with the task (Pedaste et al., 2015). Through dialogue, students can practice their reasoning abilities and develop a deeper understanding of complex concepts. However, unless they are explicitly taught to do so, students do not often share their thoughts, orient and listen to one another, deepen their reasoning, or engage with each other's reasoning (Chinn et al., 2000; Gillies, 2019; Michaels & O'Connor, 2012). As a consequence, guidance towards productive dialogue is broadly regarded as indispensable (Hu & Chen, 2023).

Technology can offer guidance as part of the collaborative inquiry learning approach (de Jong, 2019). For instance, prior studies reported successful implementations of guidance in dialogue by employing collaborative conversational agents (CCAs) during online lessons (Kumar & Rosé, 2011). CCAs are software agents designed to help student groups engage in productive dialogue by incorporating instructional strategies (Adamson et al., 2014). By identifying patterns in student dialogue, the application of learning analytics can assist automatic interventions by providing insights into students' behaviour and their contexts (Banihashem et al., 2022; Gašević et al., 2015; Siemens & Long, 2011). However, studies into the impact of CCAs on student dialogue remain limited, posing a compelling need for further research.

This paper investigates the impact of a specific CCA, Clair (Collaborative Learning Agent for Interactive Reasoning), on student dialogue using data from two classroom studies in collaborative inquiry learning settings. First, we conceptualize the problem and outline our research questions. Then, we describe how we modelled Clair's interventions, how learning analytics can help trigger them in a variety of contexts, and how we evaluated student dialogue. Finally, we discuss aspects of dialogue that improved under the influence of Clair and those that require additional attention in future work.

1.1 | Background

1.1.1 | Productive student–student dialogue in collaborative inquiry learning

By doing scientific experiments, students can explore relations between concepts, infer new knowledge, and test ideas (Chu et al., 2016), a process known as an 'inquiry cycle'. Inquiry cycles differ, but they usually include the phases of orientation, conceptualization, investigation, and conclusion, while students communicate and reflect with each other (Pedaste et al., 2015). Collaboration during inquiry cycles can lead to a valuable synergy for increased learning outcomes, yet its efficacy depends heavily on the students' ability to work together, a skill that must be cultivated rather than assumed (Cohen, 1994), and their shared commitment to achieve a collective goal (Miyake & Kirschner, 2014). Buiding on the work of others (Mercer, 2000; Rojas-Drummond et al., 2010), we view collaborative dialogue to be productive when learners critically and constructively exchange ideas, thereby making proposals, knowledge and reasoning visible. Collaborative dialogue appears to be influenced by many variables, including: specific teaching practices being used specific materials being used; time for students to express their ideas; group composition; and group discussion support (Rapanta et al., 2023).

The Academically Productive Talk (APT) framework, also referred to as 'Accountable Talk' (Michaels et al., 2008, 2016), is a widely used strategy for supporting students to engage in collaborative dialogue. According to Adamson et al. (2014), "APT is a classroom discussion facilitation approach that has grown out of instructional theories that emphasize the importance of social interaction in the development of mental processes, in particular ones that value engaging students in transactive exchanges." (p. 97). With a content-independent strategy, APT helps students direct their comments and questions to each other, thus challenging them to reason together and on their own. A few examples of APT's interventions (known as 'talk moves') include "Recapping" (i.e., "*Could someone summarize what we have discussed so far?*"), "Linking contributions" (i.e., "*Mike, how does that link to what Sofia just said?*"), and "Example" (i.e., "*Sofia, could you give us an example?*"). APT was developed out of a Vygotskian theoretical framework (Michaels et al., 2008) and has a strong consideration for the needs and characteristics of scientific dialogue (Michaels & O'Connor, 2012).

1.1.2 | Productive dialogue requires guidance

With sufficient turn-taking, dialogue can be highly productive when each person in the group constructively contributes with additional ideas that go beyond the learning material at hand (Chi & Wylie, 2014). However, this is not always the case, for example when verbal students dominate the conversation, leaving others to agree superficially without adding their thoughts (Chinn et al., 2000). Leaving the inquiry cycle entirely up to the students, particularly the novice ones, is often associated with lower learning gains, and students need guidance to fully benefit from this approach (de Jong et al., 2023). Specific guidance can be implemented through (real-time) prompting to support productive dialogue (Lazonder & Harmsen, 2016), for instance, by prompting the quieter students to share their thoughts or asking the dominant students to encourage others to contribute more. Yet, to employ such talk moves, teachers need to anticipate dialogue paths, frame thought-provoking questions, select appropriate talk formats, and strategically address student misconceptions (Michaels & O'Connor, 2012). This task becomes even more challenging when several groups need support simultaneously.

As collaborative learning can be counterproductive (Cohen, 1994; Miyake & Kirschner, 2014), CCAs have been proposed by previous studies as scalable solutions to offer guidance on student dialogue (Sikström et al., 2022). CCAs can act like members of a student group, potentially facilitating discussions and directing students' collaboration (Kumar & Rosé, 2011). A content-independent talk strategy such as APT is especially relevant as it allows CCAs to scale its usage to a diversity of concepts (Adamson et al., 2014). For instance, Tegos et al. (2016) reported positive effects on students' explicit reasoning and balanced participation for (university) students, in dyads, who interacted with an APT-based CCA. Their study also found a positive, indirect, effect of using a CCA on learning outcomes, which was mediated by the increase in explicit reasoning. Also, Nguyen (2023) reported positive effects on learning outcomes and transactive exchanges in favour of (9th-grade) students, in groups of five, who interacted with a CCA. This effect was especially prominent when the CCA was presented as a 'less-knowledgeable' peer (ibid.). Similarly, Adamson et al. (2014) investigated the effect of a CCA on learning outcomes and transactive exchanges, reporting findings from four studies with students from 9th-grade and university level, in groups of three, and found mixed results across and within audiences. In their study, the effectiveness of CCAs varied due to, among other things, the ability level of learners and the structure of the material utilized. Overall, research on CCAs in online collaborative learning is limited, especially in inquiry-based learning settings, and it is still unclear how CCAs may better support productive dialogue among students.

1.1.3 | Assessing productive dialogue and students' perceptions

To assess the effects of talk strategies on student dialogue, Michaels and O'Connor (2015) elaborated on the idea that teacher talk is productive when the "Four Goals for Productive Discussions" (FGPDs) are met. These are:

- Goal 1—"Helping individual students share their own thoughts"
- Goal 2—"Helping students orient to and listen carefully to one another"
- Goal 3—"Helping students deepen their reasoning"
- Goal 4—"Helping students engage with others' reasoning"

We argue that FGPDs provide an interesting lens to explore how CCAs impact student dialogue because they describe observable behaviours and align well with the learning goals of using CCAs in other studies, such as explicit reasoning (Tegos et al., 2016; Weinberger & Fischer, 2006), transactive exchanges (Adamson et al., 2014; Noroozi et al., 2013), etc. FGPDs can be seen as a reinterpretation and expansion of the intended effects of CCAs in dialogue; further, by encapsulating core tenets of APT, they can be used to assess the *productivity* of dialogue on multiple aspects. Within this context, learning analytics involves the measurement, collection, analysis, and reporting of data about learners and their environments to enhance our comprehension of learning and possibly also to form a basis for optimizing the environment (Banihashem et al., 2022; Gašević et al., 2015; Siemens & Long, 2011). We assert that learning analytics, by mining patterns in the dialogue (Valle Torre et al., 2023), can provide reliable data-driven estimations of how effectively students are meeting the FGPDs.

In addition to what can be observed in dialogue, the students' perception of the importance of their active participation is a key factor that characterizes productivity (Howe et al., 2019). To more concretely assess students' perceptions of the productivity of their dialogue. the three accountability dimensions proposed by Michaels and colleagues (Michaels et al., 2008; Michaels et al., 2016) can be used. According to Michaels et al., students are accountable (a) to the *learning community*, when they seek to respectfully listen to others and build responses on each other's contribution; (b) to *accurate knowledge*, when they seek to ensure gather sufficient and correct information to support their contributions; and lastly (c) to *rigorous thinking*, when they seek to make coherent and structured arguments (Michaels et al., 2016). These dimensions are particularly suitable because they directly capture the essence of students' perception of the importance of their participation in the perspective of APT. By gauging each of these dimensions (e.g., through a questionnaire), we could gain insight into students' perceived productivity.

1.2 | Research questions

Prior studies, such as those by Adamson et al. (2014), Tegos et al. (2016) and Nguyen (2023), have highlighted how CCAs impact dialogue productivity by analysing aspects such as transactive exchanges, explicit reasoning and balanced participation, demonstrating the potential of CCAs in facilitating dialogue. Yet, literature lacks further exploration of the effects of CCAs in different dimensions of dialogue productivity, as defined by the Four Goals for Productive Discussions (FGPDs). The present study seeks to extend this understanding, especially in an inquiry-based learning setting, by examining student dialogue with and without a CCA. Firstly, the following research question is explored:

> **RQ1.** To what extent does the collaborative conversational agent make written student dialogue more productive compared to when the agent is not present? More specifically, are there differences in the way students:

a. share their thoughts?
b. orient and listen to one another?
c. deepen their reasoning?
d. engage with each other's ideas?

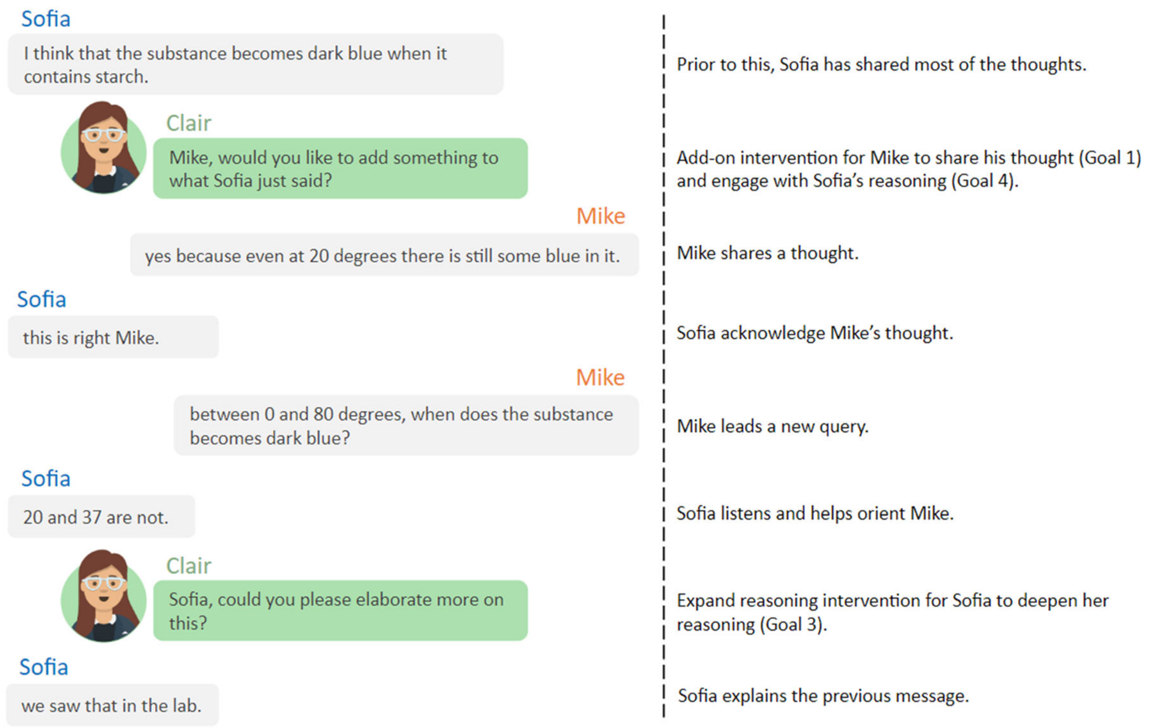

FIGURE 1 Examples of interventions from Clair in a dialogue.

Additionally, understanding how students perceive the productivity of their dialogue in the presence of CCAs complements the more observable effects. The present study aims to shed light on the potential shifts in students' perceptions by assessing three accountability dimensions, as elaborated by Michaels et al. (2008), in environments both with and without a CCA. Secondly, the following research question is explored:

> **RQ2.** To what extent does having the collaborative conversational agent present affect students' perception of their dialogue compared to when the agent is not present? Are there differences regarding:

a. learning community?
b. accurate knowledge?
c. rigorous thinking?

2 | COLLABORATIVE LEARNING AGENT FOR INTERACTIVE REASONING (CLAIR)

When designing automated interventions in student dialogue, the main design challenges concern the type of intervention needed and its timing (Kumar & Rosé, 2011). To address the former, Clair employs eight talk moves that were crafted using the APT guidelines (Michaels et al., 2016). To make informed decisions on the latter, Clair employs learning analytics to evaluate the dialogue and decides on interventions utilizing a triggering mechanism. In Figure 1, a dialogue between Sofia and Mike is depicted, where Clair encourages Mike to share his thoughts and to engage with Sofia's reasoning (Goal 1 and Goal 4), and attempts to help Sofia deepen her reasoning (Goal 3). In this section, we describe the specific talk moves employed and then dive into the process that Clair follows to trigger APT interventions.

2.1 | Talk moves

Table 1 presents the eight talk moves employed by Clair. For instance, the "Recapping" move prompts students to articulate the ideas they have discussed so far, while the "Rephrasing" move prompts students to be on the same page, promoting shared understanding. Apart from the Recapping move, Clair's talk moves include directionality (Tegos et al., 2016) by addressing either the speaker (i.e., the last student who spoke) or the discussant.

Each talk move serves a unique role and the actual impact of Clair depends on which ones are triggered. For example, if one student is providing more topic-related contributions, talk moves addressing the discussant are much more likely to happen. Overall, by employing various talk moves, Clair aims to comprehensively support students in the FGPDs based on what happens in the dialogue. To make interventions sound more natural and less repetitive, Clair employs three distinct variants of each talk move maintaining the same meaning. Figure 2 illustrates the overall process, further explained on the following subsections, for triggering talk moves in student dialogue.

TABLE 1 Description of talk moves used in this work.

Talk moves	Description	Examples of what Clair says
Recapping	Summarizing key ideas to understand the progress of the dialogue.	"This conversation is interesting. Would any of you be able to give a brief summary of what you've covered so far?"
Add-on	Building on the ideas of others by adding new information or perspectives.	"Mike, would you like to add something to what Sofia just said?"
Rephrasing	Restating the ideas of others in different words ensures that everyone understands what has been said.	"Sofia, could you put in other words what Mike just said?"
Agree/Disagree	Expressing agreement or disagreement with the ideas of others.	"Mike, do you agree or disagree with Sofia?"
Linking contributions	Connecting the ideas of others to the main topic or question being discussed.	"Sofia, how does your ideas align with what Mike just said?"
Build on prior knowledge	Connecting new information or ideas to what students already know.	"Mike, how does this expand what Sofia have said so far?"
Example	Providing an example or illustration to help explain a concept or idea.	"Sofia, could you give an example?"
Expand reasoning	Asking students to explain why they believe something, and how they arrived at a certain conclusion.	"Mike, could you please elaborate more on this?"

2.2 | Dialogue variables

To help identify moments to intervene, Clair uses twelve dialogue variables as a set of learning analytics instruments, which can be interpreted as sensors for each dialogue message. The dialogue variables are utilized in Clair's pattern recognition component (Table 2). Like the talk moves, the dialogue variables are content-independent to allow higher transferability of the pattern recognition component to different topics.

In a previous paper, we proposed and evaluated ConSent (de Araujo et al., 2023b), a machine-learning algorithm powered by the pre-trained multilingual Universal Sentence Encoder (mUSE, Yang et al., 2019) to perform automated content analysis of dialogue messages. In the current work, ConSent leverages eight dialogue variables related to measuring the message's 'focus' and 'intent' (see Table 2). ConSent models have achieved substantial reliability when compared to human raters ($\kappa = 0.60$ for 'focus' and $\kappa = 0.63$ for 'intent', on chats in the topic of electric circuits, in Dutch). Also, it has been found that these models can be transferred, albeit with a drop to moderate reliability to dialogue messages on different topics and in a different language ($\kappa = 0.44$ for 'focus' and $\kappa = 0.52$ for 'intent', on chats in the topic of radiation, in Portuguese). Moreover, these models seem to be reliable in languages that mUSE handles that were not present in ConSent's training data (de Araujo et al., 2023b), which further increases the usability of Clair to the learners' native language.

In addition to the eight ConSent variables aimed at identifying the 'focus' and 'intent' of messages, four additional variables were used to describe the student dialogue more completely. Topic similarity identifyies similarity to topic keywords, topic accumulation indicaes the proportion of the current speaker's contributions regarding this similarity, time spent refers to time spent since the chat started, and messaging speed indicates pace in terms of the number of messages per minute. The combination of these twelve dialogue variables helps Clair to decide when to intervene.

2.3 | Triggering mechanism

Clair presents a talk move after the triggering mechanism is activated. Based on the values of the dialogue variables of each message, Clair's triggering mechanism implements a fuzzy expert system that relies on rules to associate target states of dialogue variables (inputs) with triggering states of talk moves (outputs). We opted to employ a fuzzy expert system as it strikes a balance between interpretability and the ability to handle uncertain and imprecise data (Zadeh, 1983). Our implementation was developed using the Python library Scikit-Fuzzy (Warner et al., 2019). Further explanations of the steps required for utilizing fuzzy logic in the context of learning analytics are given by Casalino et al. (2022).

In our fuzzy expert system, each dialogue variable was modelled with three fuzzy sets that describe its intensity as *high*, *medium*, or *low*. One advantage of this approach is that this allows for logical operations to be made on ranges rather than using static thresholds, thereby helping to deal with uncertainty and imprecise variables (Zadeh, 1983). The specific ranges for each output are determined by the fuzzy expert system given the message being processed, as well as by the rule base. In particular, the rules follow the structure of $IF\ (x_1\ is\ A_1)\ and...(x_n\ is\ A_n)\ THEN\ (y_1\ is\ B_1)\ and...(y_m\ is\ B_m)$, where A_i and B_j are fuzzy sets to describe the x_i dialogue variable and y_j talk moves respectively. One example of a rule is "IF (DOM is *high*) and (TSIM is *high*) and (TACC is *low*) THEN (ADD_ON is *active*)", which corresponds to the condition where the message is focused on the domain, mentions words close to the topic keywords, and the speaker has mentioned keywords less often than the discussant. For each dialogue message, the fuzzy expert system processes the degree to which this rule was met and finds a final value of *active* for that talk move.

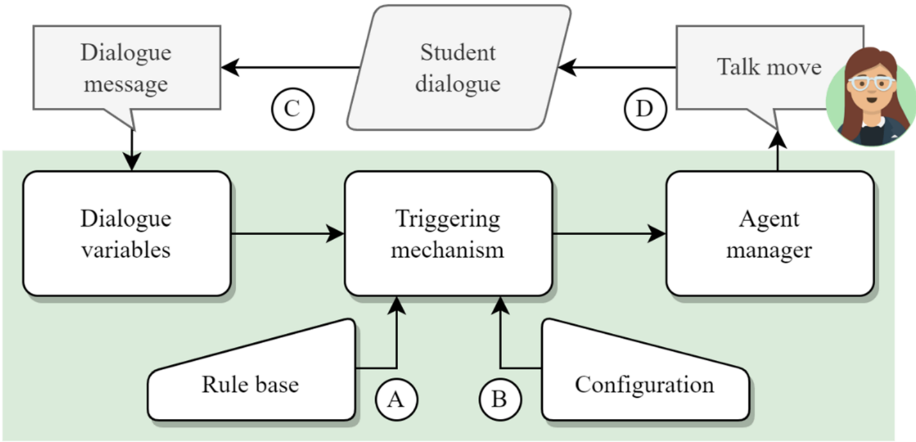

FIGURE 2 Flowchart of Clair's internal components. When setting up, Clair's triggering mechanism must be prepared with a general rule base (a), and then with a configuration specific to the task at hand, including topic keywords (b). After that, Clair can start receiving messages (c) and send talk moves back to the dialogue (d).

To ensure that the rule base is simple to interpret and adjust, we currently employ one rule to determine when the talk move *active's* output should be *high* (i.e., a sensitive dialogue message), and another to determine when it should be *low*, (i.e., a dialogue message not to intervene). The formulation of rules and subsequently preliminary evaluation of Clair's APT interventions was accomplished by querying secondary school teachers (de Araujo et al., 2023a).

2.4 | Agent manager

The agent manager's role is to prioritize and select the most appropriate talk move to avoid repetition and enhance the student experience. First, the agent manager selects the main talk move candidate based on the final active values explained above. Second, if a candidate talk move has been used within the last intervention, it is not considered. Third, also to prevent repetition, the frequency of talk move usage is tracked so that, when two talk moves have the same active value and the same usage frequency, the least used talk move is selected. In case two talk moves are tied, the agent manager selects one randomly. Once a talk move is selected, an utterance variation is chosen randomly. By employing these criteria, the agent manager adds more structure to the outputs produced from the triggering mechanism, ultimately enhancing the student experience.

3 | METHOD

3.1 | Experimental design

We conducted two studies with a consistent experimental design across both studies, ensuring a certain comparability of results between the two distinct contexts. In the experimental design, there were two phases. The first phase served as a baseline where dyads engaged in the learning task and discussed it in the chat without Clair's presence. In the second phase, only the treatment dyads had Clair, allowing us to control for spontaneous improvements in dialogue quality over time.

Before the first phase, we randomly assigned students to dyads. Next, dyads were assigned to control (Clair absent) or treatment (Clair present) conditions after the first phase, based on stratified sampling. Specifically, we fed the first-phase dialogues to Clair and we ranked all dyads based on the number of interventions they would have received. We then assigned them to the two conditions alternately based on the ranking. This was done to ensure that the starting range of dialogue productivity was similar in the treatment and control conditions.

3.2 | Participants

Both studies took place during science classes in secondary schools. Since the studies were conducted as part of the schools' normal curricula, passive consent was applied. All personal information gathered was anonymized after the studies ended. The research was approved by the Ethics Committee of the University of Twente.

In Study 1, 60 students from two 8th-grade biology classes at a private high school in Americana, Brazil, were enrolled. Half of the students were assigned to the control condition and half to the treatment condition. Only dyads that attended all sessions were included in the analysis. The final sample consisted of 19 dyads: 9 treatment and 10 control; 18 female and 20 male; 12–14 years old.

In Study 2, 100 students from five 9th-grade biology classes of two public high schools in Tubbergen and Almelo, the Netherlands, were enrolled. In Study 1, we observed that a few treatment dyads effectively shifted to the control condition, receiving zero (or close to zero) interventions during their dialogue. Therefore, in Study 2, treatment dyads with less than two Clair's interventions were not considered in the analysis. To still have sufficient dyads in the experimental condition, one-third of the dyads were assigned as control and the other two-thirds as treatment. The final sample, after removing the dyads that received less than two interventions consisted of

TABLE 2 Description and examples of the dialogue variables.

Dialogue variable	Description	Examples
ConSent—Focus		
Domain (DOM)	Probability that the message is a domain-related utterance, that is, discussing domain-related topics, responding to questions or domain-related information, or encouraging others to look at domain-related content.	"Parallel circuits are different"; "I think the answer is simple"; "Where you see the definitions?"
Coordination (COO)	Probability that the message is discussing HOW to complete the learning tasks, discussing who does what, where they are done, or where to go in the learning environment (not domain related, but can be task-related.)	"We should start question 1"; "I need to check again".
Off-task (OFF)	Probability that the message is discussing something unrelated to the domain or task, or responding to something unrelated to the domain or task.	"Hello there!"; "I am fine and you?"; "let's play a game".
ConSent—Intent		
Informative (IN)	Probability that the message is an informative utterance/response about knowledge/facts, actions or personal state.	"I don't know the answer"; "Ok!"; "the result is 3 V"; "If it's blue then it's done".
Argumentative (AR)	Probability that the message is an argumentative utterance/response aiming at the following: clarifying, reasoning, interpreting, making a critical remark, stating conditions, drawing conclusions using words such as "but", "because", "due to", "in my opinion" etc.	"This happens because of the parallel circuit"; "In my opinion, I think that is a good answer, but I am not sure".
Asking for information (AI)	Probability that the message is asking for understanding, explanations, clarification or assistance.	"What are parallel circuits?" "Can you help me?".
Active Motivating (AM)	Probability that the message is encouraging (a) group member(s) to participate or to take action.	"C'mon Laura, let's do it"; "we are doing great!".
None of these (NOS)	Probability that the message is none of the other content codes.	Incomplete sentences, chitchat, or unclear intent.
Additional dialogue variables		
Topic similarity (TSIM)	Semantic similarity between the message's content and pre-configured topic keywords. More specifically, it computes the inner product of the message's embeddings *x* and the keyword's embeddings *k*, assuming that the embeddings are normalized.	High: "Parallel circuits are different of serial in current"; Medium: "I think it should be in parallel"; Low: "I found the answer."
Topic accumulation (TACC)	Speaker's proportion of topic similarity accumulated thus far relative to the total, including the discussant.	High: speaker contributed more than discussant; Medium: speaker and discussant are equitable; Low: speaker contributed less than discussant.
Time spent (TIME)	Time in seconds since the activity has started.	High: 30 min; Medium: 15 min; Low: 5 min.
Messaging speed (PACE)	Ratio of messages by unit of time, measured in messages per minute.	High: 30 messages/min; Medium: 10 messages/min Low: 5 messages/min.

28 dyads: 13 control and 15 treatment; 29 female and 27 male; 14–16 years old.

3.3 | Materials

In both studies, all the materials, including Clair's talk moves, were prepared in the student's native language (i.e., Portuguese in Study 1 and Dutch in Study 2) and delivered through Go-Lab,[1] an online learning ecosystem that provides access to virtual labs, learning resources, and (collaboration) tools (de Jong et al., 2021). In Go-Lab, teachers can create, publish, and use Inquiry Learning Spaces (ILSs). An ILS combines an online lab with multimedia material and learning tools (apps). Students can navigate through different phases in the ILS, interact with a science lab, and complete questionnaires while also collaborating with partners.

Clair was integrated into Go-Lab's chat app which students had access to during the whole task. One important difference between studies was that Clair's rule base was further refined before Study 2. Specifically, this refinement was focused on reducing Clair's

[1] https://www.golabz.eu/

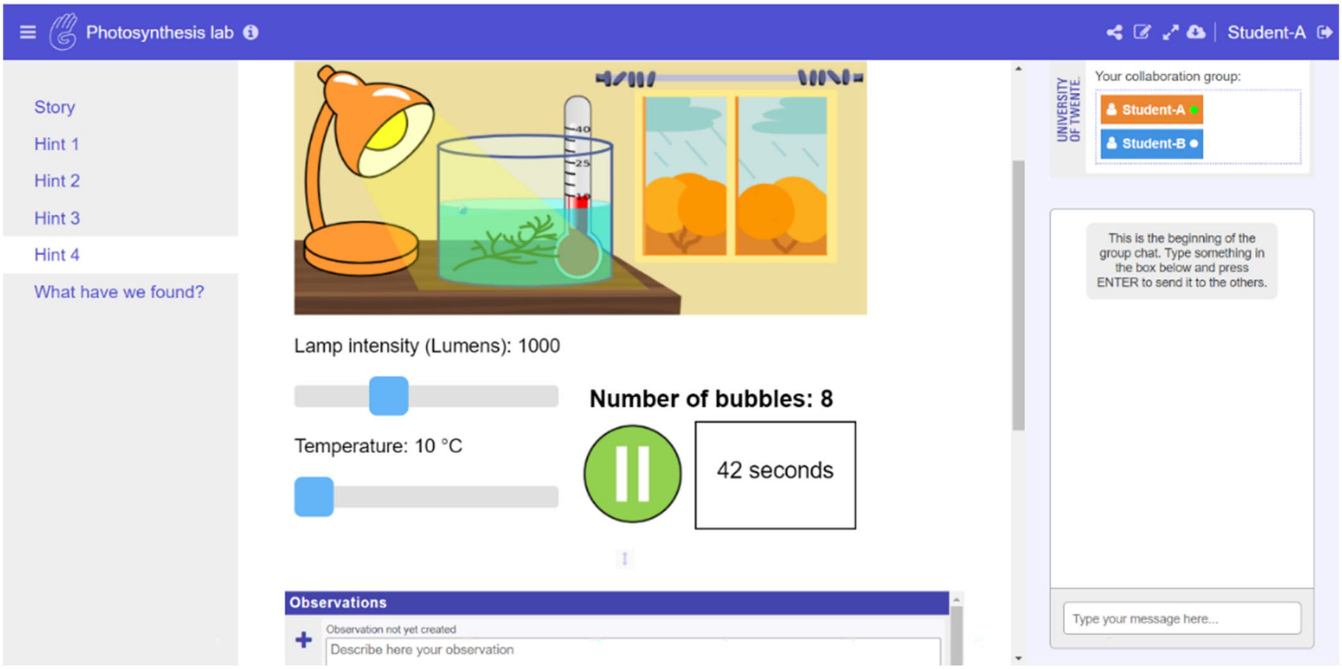

FIGURE 3 Screenshot from photosynthesis ILS.

repetitiveness and timing, an issue found after a round of expert evaluation with the teachers from Study 1 (de Araujo et al., 2023a).

Both ILSs used in Study 1 and Study 2 were co-designed with the teachers to ensure that the content was relevant for their students. The ILSs were carefully designed to require constant peer communication to complete the task successfully. We incorporated prompts within the ILSs to stimulate the exchange of information such that genuine dialogue would take place. Both control and treatment dyads had access to the same ILS with Clair's presence within the chat being the only difference.

To measure students' perceptions of dialogue productivity, we used the student-perceived productivity (SPP) questionnaire, an adaptation of the students' discursive engagement questionnaire from Chen et al. (2020). In the SPP questionnaire, the dimensions are scales, calculated by averaging the responses to a set of three questionnaire items focused on each of the dimensions (see Table 5). In both studies, we applied the SPP questionnaire using a 5-point Likert scale.

3.3.1 | Study 1—Photosynthesis ILS

The photosynthesis ILS aimed to help students understand the role of temperature and light exposure in the photosynthesis process and was designed to last four lessons. In the task, students investigated what happens to plants in different seasons and were asked "To what extent do temperature and light exposure influence the photosynthesis process?" Students had access to an interactive simulation in which they experimented with temperature and light intensity (see Figure 3), collected data, and discussed the results. Throughout the task, students were constantly prompted to discuss with their peers, sharing their hypotheses, findings, and conclusions. This helped students interact more often across the whole task. Towards the end, the task required students to reflect on the initial question and their initial and final thoughts.

3.3.2 | Study 2—Enzymes ILS

The enzymes ILS aimed to help students understand the role of enzymes in the digestive system and was designed to last two lessons. In the task, students first read the content and were asked to remember the names and functions of enzymes in the digestive system. Then, students had to formulate and discuss hypotheses about how temperature influences how saliva breaks down starch. After that, students interacted with an online lab where they observed what happens when saliva and starch are mixed at various temperatures (see Figure 4). Students collected data and analysed results. Similar to the photosynthesis lab task, students were constantly encouraged to discuss with their peers throughout the whole task. In the end, the task required students to reflect on their steps and what they had discovered during the lesson.

3.4 | Procedure

A similar procedure with local adaptations was used in both studies. Study 1 spanned four lessons, each lasting 40 min, while Study 2 spanned two lessons, each lasting 50 min. At the start of the first lesson, the task and the supporting role of Clair in it were verbally

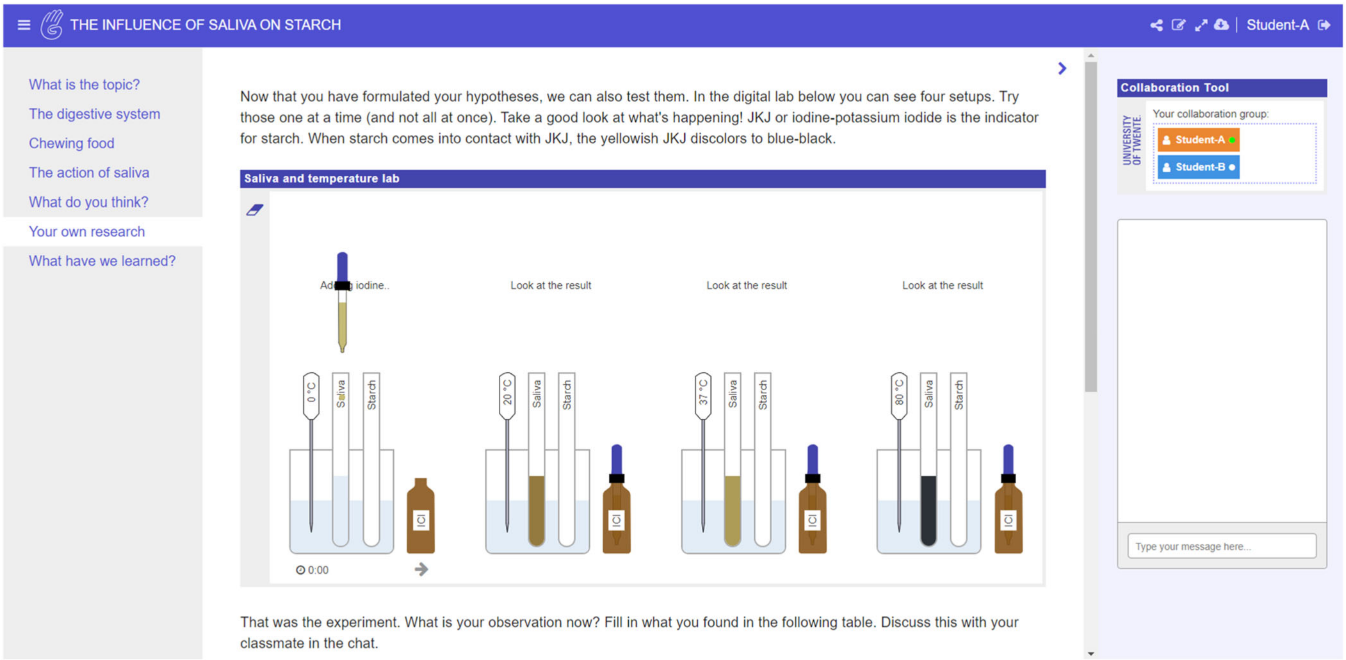

FIGURE 4 Screenshot from the enzymes ILS.

explained. During this short explanation, students were advised that they should try and engage with Clair's interventions, but they could ignore them if they would find them not useful. Students then received the link to log in to the ILS.

Students were instructed to only discuss through chat and the teacher emphasized the importance of collaborating to progress in the task. Students who were grouped together were told to take seats apart from each other. This is similar to a situation where collaboration takes place remotely. At the start of the second phase (i.e., the third lesson in Study 1 and the second lesson in Study 2), students were notified that Clair was present in the chat for some of the dyads. For Study 1, which had a more extended timeframe, the SPP questionnaire was made available 10 min before the conclusion of both Phase 1 and Phase 2 (i.e., second and fourth lessons). As Study 2 had a more condensed format, the questionnaire was made available 10 min before the end of Phase 2 only (i.e., the second lesson) to avoid consuming extra time that could be needed for all students to complete the task.

3.5 | Data analysis

In analysing RQ1, we adopted a learning analytics approach, utilizing the ConSent dialogue variables—namely 'focus' and 'intent' (refer to Table 2)—to operationalize measurements of the FGPDs. Within the realm of learning analytics, sequential pattern mining is a method to trace target student behaviours in data (Valle Torre et al., 2023). By leveraging this method, we determined the outcome measures for each dialogue based on the frequency of specific sequential 'patterns', or periodic trends theorized to represent behaviours associated with the FGPDs in the dialogue.

The search for the FGPDs' patterns happens within moving-windowed sequences of n dialogue messages (where n is the window size). Specifically, a pattern condition is a series of three steps, each establishing one condition assessed on individual messages within the sequence. These conditions do not necessarily need to match in three adjacent messages since chat communication is often highly dynamic, for example, with several messages being typo corrections, brief acknowledgments, or shallow responses. Some goals have patterns that often appear across shortly consecutive messages, such as Goal 1 and Goal 2, but others, such as Goal 3 and Goal 4, have patterns appearing in chat on larger windows, in between comparatively more peripheral intermediary remarks. After inspection of $n = 5$ and $n = 7$, in order to capture patterns across all goals with a consistent parameter, we identified that $n = 7$ would convey a dialogue mini-episode which sufficiently encompasses the three messages that could meet the conditions, plus four others that could be peripheral remarks. Figure 5 illustrates the moving-window segmentation approach for counting pattern matches given a single goal.

More specifically, Table 3 presents the patterns for each of the FGPDs. For instance, the pattern defined for Goal 1 (i.e., sharing thoughts) requires that, within the sequence dialogue messages, a student posted a task-related informative statement in the chat, then followed this up with an argumentative statement, while on the third step, either one of the dyad students posted a task-related informative or argumentative statement. In other words, the pattern for Goal 1 tries to identify sequences in which students *shared their thoughts* by examining task-related informative and argumentative statements. Goal 2's pattern helps identify when students were answering each

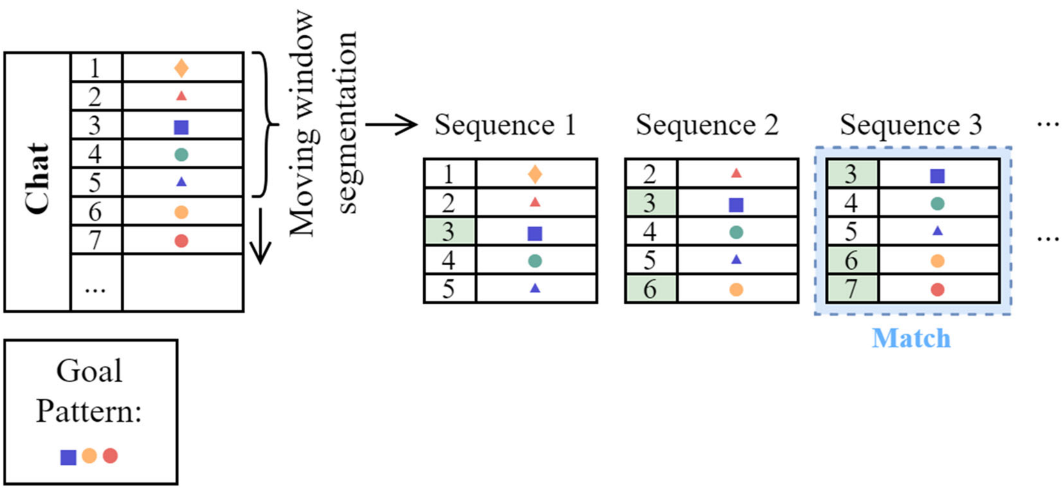

FIGURE 5 Example of pattern mining with the moving window segmentation approach with $n = 5$. Given a defined goal pattern that takes into account certain conditions based on the dialogue variables, the approach counts the sequences of dialogue messages that contain the pattern.

TABLE 3 Patterns condition steps for each of the goals including a real example of dialogue messages of a possible match.

Goal	Step	Pattern condition	Example
Goal 1—Students share their thoughts	1	OFF ≤ *low* and IN ≥ *high coming from any speaker*	Student-1: yes, because at 20 degrees there is still some blue black visible
	2	OFF ≤ *low* and AR ≥ *medium* coming from the same speaker of step 1	Student-1: And at 37 degrees it is even lighter again
	3	OFF ≤ *low* and (IN ≥ *high* or AR ≥ *medium*) *coming from a different speaker*	Student-2: at 0 and 80 degrees it is blue-black
Goal 2—Students orient and listen to one another	1	OFF ≤ *low* and AI ≥ *high coming from any speaker*	Student-1: what to do?
	2	OFF ≤ *low* and (IN ≥ *high* or AR ≥ *medium*) coming from a different speaker	Student-2: the place my arrow is you don't have to fill in anything
	3	OFF ≤ *low* and (IN ≥ *high* or AR ≥ *medium*) *coming from the same speaker of step 2*	Student-2: only under here it says what you have to fill in
Goal 3—Students deepen their reasoning	1	DOM ≥ *medium* and (IN ≥ *high* or AR ≥ *medium*) coming from any speaker	Student-2: I dont know exactly
	2	DOM ≥ *medium* and (IN ≥ *high* or AR ≥ *medium*) coming from the same speaker of step 1	Student-2: the previous one instead of little starch it has no starch
	3	DOM ≥ *medium* and (IN ≥ *high* or AR ≥ *medium*) coming from the same speaker of step 1	Student-2: because otherwise it would have to be a bit darker
Goal 4—Students engage with other's reasoning	1	DOM ≥ *medium* and (AR ≥ *medium*) coming from any speaker	Student-2: because otherwise it would have to be a bit darker
	2	DOM ≥ *medium* and (AR ≥ medium or AI ≥ *high*) coming from a different speaker	Student-1: Also think that's the right answer
	3	DOM ≥ *medium* and (IN ≥ *high* or AR ≥ *medium*) coming from the same speaker of step 1	Student-2: yes, but not very dark

Note: The pattern conditions rely on ConSent dialogue variables regarding 'focus', namely, Domain (DOM), Off-task (OFF), and 'intent', namely, Informative (IN), Argumentative (AR), and Asking for information (AI), plus information about who is the message's speaker.

other's questions with contributions related to the task, thereby *orienting and listening carefully to one another*. Goal 3's pattern helps identify when students were extending their own task-related contributions, thereby *deepening their reasoning*. Lastly, Goal 4's pattern helps identify when students were discussing an argument together, thereby *engaging with each other's reasoning*. For each of the

TABLE 4 Average counts of pattern matches for all goals on each study per phase and per condition.

Goal	Study 1				Study 2			
	Phase 1 M (SD)		Phase 2 M (SD)		Phase 1 M (SD)		Phase 2 M (SD)	
	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
Goal 1 “Share”	8.300 (7.732)	9.777 (4.684)	12.200 (7.857)	18.111 (6.050)	14.615 (9.097)	12.533 (9.403)	17.231 (12.125)	21.467 (11.262)
Goal 2 “Listen”	3.700 (2.983)	3.889 (2.848)	6.100 (4.677)	8.333 (5.172)	5.077 (3.353)	5.000 (3.854)	5.000 (1.78)	7.000 (3.78)
Goal 3 “Deepen”	1.700 (2.791)	2.556 (2.963)	2.200 (2.974)	3.667 (4.183)	1.923 (2.216)	1.933 (2.12)	4.923 (4.752)	5.933 (6.296)
Goal 4 “Engage”	2.200 (2.936)	3.778 (3.598)	3.100 (3.107)	3.778 (2.774)	3.308 (2.898)	2.000 (1.813)	5.692 (5.588)	5.733 (6.147)

goals, their outcome measure is the count of unique pattern matches found.

To answer RQ1, the same statistical comparisons were conducted for both studies using the outcome measures for each of the goals. Violations of the assumptions underlying parametric tests, mainly regarding normality, linearity, and homogeneity of regression slopes, were found among all outcome measures, thus non-parametric tests were conducted throughout. First, non-parametric ANCOVA, with the outcome of Phase 1 as a covariate was employed to analyse differences between conditions in Phase 2 while accounting for any initial disparities. Second, Wilcoxon's signed-rank tests were independently conducted for each condition to detect significant shifts in outcomes from Phase 1 to Phase 2. Employing both tests was important to capture both between-condition differences and within-condition changes over time. In addition, to check whether dyads are comparable prior to comparing the outcomes of Phase 2, for each goal, the Mann–Whitney U test was conducted to assess differences between conditions in Phase 1's outcome. All statistical tests were performed at a 0.05 level of significance.

To answer RQ2, we needed to employ different statistical approaches for each study. For Study 1, as data is available for both phases, we applied the same statistical tests as in RQ1. For Study 2, as the questionnaire was measured only at the end of Phase 2, conditions were then compared using the Mann–Whitney U test in order to determine if the presence of Clair was associated with differences between conditions in students' perceived productivity.

4 | RESULTS

> **RQ1.** To what extent does the collaborative conversational agent make written student dialogue more productive compared to when the agent is not present?

Table 4 summarizes the average pattern matches of all goals. For both studies, Mann–Whitney U tests showed no significant differences in Phase 1 between the control and treatment dyads (i.e., C_{ph1} vs. T_{ph1}, where C_{ph1} refers to the control condition in Phase 1) ($p > 0.05$), thus ensuring that any differences observed in Phase 2 between conditions were not influenced by pre-existing differences in Phase 1.

For Goal 1, non-parametric ANCOVA showed no significant differences between the control and treatment dyads in Phase 2 (i.e., C_{ph2} vs. T_{ph2}) ($p > 0.05$). This was true for both studies. Regarding dyads' activity from Phase 1 to Phase 2, Wilcoxon's signed-rank test showed that the count of identified patterns increased significantly for the treatment condition (i.e., T_{ph1} vs. T_{ph2}) in both studies (Study 1: $Z = 2.380$, $p = 0.017$, $r = 0.793$; Study 2: $Z = 2.578$, $p = 0.001$, $r = 0.665$). On the contrary, no significant difference was observed for the control condition (i.e., C_{ph1} vs. C_{h2}) of either study ($p > 0.05$).

For Goal 2, non-parametric ANCOVA showed no significant differences between the control and treatment dyads in Phase 2 ($p > 0.05$). Wilcoxon's signed-rank test also showed no significant

TABLE 5 SPP questionnaire scales, items, and their average response.

Scale	Questions	Study 1 M (SD)	Study 2 M (SD)
Learning community	I listened to my classmate without interrupting until it was my turn to speak.	4.237 (0.764)	4.509 (0.669)
	I listened to my classmate's opinions to get inspiration.	4.276 (0.873)	4.717 (0.495)
	I defended my position respectfully while discussing with my classmate.	4.211 (0.853)	4.151 (0.864)
Accurate knowledge	I gathered information to support my ideas.	4.408 (0.803)	4.192 (0.687)
	I checked whether the information that our group gathered was correct.	4.092 (0.955)	3.830 (0.871)
	I discussed with my classmate what information was needed to progress on the task	4.118 (0.923)	4.453 (0.695)
Rigorous thinking	I checked whether our arguments were clear and coherent.	4.250 (0.751)	3.660 (1.018)
	I discussed whether our arguments are sufficiently convincing.	3.816 (0.905)	3.830 (0.955)
	I discussed with my classmate cases where our arguments may not be correct.	3.553 (1.259)	3.788 (0.936)

differences regarding dyads' activity from Phase 1 to Phase 2 (i.e., C_{ph1} vs. C_{ph2}, and T_{ph1} vs. T_{ph2}) of either study ($p > 0.05$).

For Goal 3, non-parametric ANCOVA showed no significant differences between the control and treatment dyads in Phase 2 ($p > 0.05$). For Study 1, Wilcoxon's signed-rank test also showed no significant differences regarding dyads' activity from Phase 1 to Phase 2 ($p > 0.05$). For Study 2, Wilcoxon's signed-rank test showed significant differences between phases on both conditions (control: $Z = 2.502$, $p = 0.012$, $r = 0.694$; treatment: $Z = 2.677$, $p = 0.007$, $r = 0.691$).

For Goal 4, non-parametric ANCOVA showed no significant differences between the control and treatment dyads in Phase 2 ($p > 0.05$). For Study 1, Wilcoxon's signed-rank test also showed no significant differences regarding dyads' activity from Phase 1 to Phase ($p > 0.05$). For Study 2, Wilcoxon's signed-rank test showed that the count of identified patterns increased significantly for the treatment condition ($Z = 2.748$, $p = 0.006$, $r = 0.709$), while no significant difference was observed for the control condition ($p > 0.05$).

Therefore, given the exclusive significant differences observed between Phase 1 and Phase 2 in Goal 1 (across both studies) and in Goal 4 (only in Study 2) for the treatment condition, it can be confirmed that Clair's presence partly contributed to more productive dialogue.

> **RQ2.** To what extent does having the collaborative conversational agent present affect students' perception of their dialogue compared to when the agent is not present?

Table 5 presents the average responses of all the questionnaire items per study. The overall reliability of the SPP questionnaire was assessed using Cronbach's alpha, found to be $\alpha = 0.889$ in Study 1 and $\alpha = 0.693$ in Study 2, indicating good and acceptable levels of internal consistency, respectively.

In Study 1, on average, the scale of rigorous thinking ($M = 3.87$, $SD = 1.03$) was perceived as much lower than those of learning community ($M = 4.24$, $SD = 0.83$) and accurate knowledge ($M = 4.21$, $SD = 0.90$). Statistical comparisons with non-parametric ANCOVA and Wilcoxon's signed rank test showed no significant differences between the control and treatment dyads and between Phase 1 and Phase 2 ($p > 0.05$).

In Study 2, once again, on average the scale of rigorous thinking ($M = 3.76$, $SD = 0.97$) was perceived as the lowest. The scale of learning community ($M = 4.46$, $SD = 0.73$) was the highest again, followed by accurate knowledge ($M = 4.16$, $SD = 0.79$), revealing a consistent outcome in students' perceived productivity across studies. Since in Study 2 the questionnaire was applied only at the end of Phase 2, unlike in Study 1, Mann–Whitney U test was used instead but showed no significant differences between the control and treatment dyads ($p > 0.05$).

5 | DISCUSSION

This work investigated the impact of Clair on productive dialogue and student-perceived productivity. Two studies were conducted with secondary school students from Brazil and the Netherlands in the context of (online) collaborative inquiry learning. While differing in some aspects across studies, the findings help to understand potential benefits when introducing Clair into students' written dialogue, as well as to discover directions for improvement in future CCA designs.

5.1 | Effects on productive dialogue

Our data clearly show that a conversational agent such as Clair can help to raise the productivity of dialogue by increasing the frequency of students “sharing their thoughts” (i.e., Goal 1 of the FGPDs framework; Michaels et al., 2016) across studies. This result is in line with Tegos et al. (2016) and Adamson et al. (2014), who suggested that APT-based CCAs can increase students' exchange of task-related contributions.

Though exclusively in Study 2, we also found differences in the frequency of students “engaging with each other's reasoning” (i.e., Goal 4) when Clair was present. In fact, we observed a close similarity in pattern averages in Phase 2 between conditions (see Table 4). The initially lower average in the treatment condition contributed to the significant difference to appear. But also, this initial disparity suggests that Clair's presence was particularly beneficial for students starting from a less favourable position, enabling them to improve. Clair may have worked differently in the two studies for this goal because of the different triggering rules adopted in Study 2, as explained in Section 3.3, while also because Study 2 was on a different topic (i.e., enzymes), had different demographics, and used a different build-up of the lessons.

Overall, these results confirm that an APT-based CCA such as Clair might help students to more frequently exchange thoughts and engage with the ideas of others, partly contributing to make dialogues more productive. In this context, we could have expected also an increase in Goal 2, which refers to students “orienting and listening to one another”, but we did not find this clearly in our data. Also, we did not find evidence that Clair helped students “deepen their reasoning” (i.e., Goal 3). The frequency of activities related to Goal 3, increased from Phase 1 to Phase 2 only for Study 2, but for the treatment and control conditions alike. This could mean that over time students engaged in deeper reasoning but that a CCA was not necessary to make this process happen. Still, as a complex endeavour with many decisions to be taken on the way, improvements can be made in each and every step, despite this we see encouraging results in two very practical goals of productive dialogue.

5.2 | Effect on students' perceived productivity

Despite the significant differences in the frequency of students “sharing their thoughts” and “engaging with each other's reasoning”, RQ2's results show that Clair's presence did not yield significant differences in student-perceived productivity. Several factors could help explain this outcome. First, Clair's interventions might have translated into implicit changes in students' perceptions that were not fully captured through the questionnaire. Second, students may have been frustrated with the guidance from the artificial agent since, after manual inspection of the data, we noticed that students have ignored Clair (too) often because the timing of interventions sometimes could have been more adequate. Assuming that teacher interventions would be successful in enhancing students' perceived productivity, the explicit non-human nature of Clair (e.g., as an artificial agent absent of personal judgement and emotional expression) might have also contributed to the lack of differences between conditions.

5.3 | Implications

Overall, our findings bring significant contributions to the use of CCAs to support productive dialogue in collaborative inquiry learning. In terms of theory-building, these findings suggest that the presence of CCAs, such as Clair (with its current set of talk moves), can enhance the frequency of students “sharing their thoughts” and “engaging with each other's reasoning”—key goals of productive dialogue. Conversely, the observation that Clair's presence was not associated with helping students “orient and listen to one another” and “deepen their reasoning” marks a turning point for theory-updating, indicating that additional talk moves are further required. Also, our research also sheds light on the complexity of the impact of CCAs, as evidenced by the lack of significant changes in student-perceived productivity despite Clair's presence.

Therefore, our findings offer insights for educators and those involved in CCA development. For educators, the insights suggest that currently these tools serve best as complementary aids, enriching rather than replacing traditional methods of guiding productive student dialogue. For researchers and CCA developers, these findings prompt a rethinking of how CCAs can be designed or used to promote students' productivity and stress the importance of multidimensional evaluations when assessing their impacts in the dialogue.

5.4 | Limitations and future work

While this work provides insights into the unique contribution of Clair towards productive dialogue and students' perceptions of productivity, we are cautious about generalizing the findings due to its limitations. First, our studies had limited sample sizes as dyads that did not attend all sessions were taken out. Comparisons between conditions with non-parametric ANCOVA using the outcomes from Phase 1 as covariate did not reveal significant differences which may have been caused by a power issue, since differences were observed in the descriptive data and in the Wilcoxon's signed rank tests. Furthermore, while the sequential pattern mining approach for goal patterns offers promising insights, its alignment with human interpretations remains a topic for future exploration, opening avenues for further validation and refinement. As a more practical issue, after manual inspection of the dialogues, we noticed that some students might have ignored Clair (too) often. This could have been because Clair's talk moves requested students to exert additional effort that they otherwise would not have to give if Clair were absent. In this case, future work could implement CCAs that are more attractive for students to use, for example, including praise or dynamic follow-up interventions. But also, Clair's timing was slightly off or another talk move would have been more adequate. The inclusion of other relevant dialogue variables as well as refinements on the triggering rules could potentially improve these issues.

Another limitation of this study is that student dialogue had to occur exclusively through chat, which limits the generalizability of our findings to more authentic collaboration settings, such as face-to-face discussions. Our results might only be warranted when multiple student dialogues are genuinely developed in the chat throughout the learning activity. Further studies of CCAs intervening on face-to-face settings might be an interesting research venue to support students in their dialogue and teachers in orchestrating the classroom.

Moreover, by investigating the impact of Clair on the four goals of productive discussions (FGPDs, Michaels & O'Connor, 2015), our findings shed light on specific goals to help future designs of CCAs for collaborative inquiry learning. Our findings underscore that future CCAs should also focus on helping students more frequently "orient and listen to one another" (Goal 2) as well as "deepen their reasoning" (Goal 3) during their dialogue. Specifically for Goal 2, future work could include more talk moves to prompt students to regularly check if they are on the same page, for example, by asking each other questions and rephrasing each other's contributions (Michaels & O'Connor, 2015). Also, talk moves can be refined to let students answer one another directly, rather than requiring an answer to the agent, as the agent's purpose is to guide student exchanges. Specifically for Goal 3, content-dependent talk moves could be employed more often in future implementations. APT guidelines include various other talk moves, including some addressing content more directly than the ones utilized in our studies, which could potentially be operationalized by the next generation of CCAs, for example, using large language models.

6 | CONCLUSION

This paper provides evidence from two classroom studies on the impact of Clair, our newly proposed collaborative conversational agent (CCA), on improving the productivity of student dialogue, and perceived productivity in collaborative inquiry learning settings. The design and analysis of Clair's interventions were guided by the core tenets of Academically Productive Talk (APT), a framework to implement teacher guidance on student dialogue. The application of learning analytics assisted in implementing Clair's triggers and part of the analysis approach. Our findings reveal that across the two studies conducted, Clair's presence resulted in a higher frequency of students sharing their thoughts compared to control dyads. In our second study, we also found that Clair's presence resulted in a higher frequency of students engaging with each other's reasoning. Conversely, no differences were found in students' perceived productivity. Future studies could utilize this knowledge to design CCAs that help students "orient and listen to one another", "deepen their reasoning", and adopt a metacognitive perspective in which students perceive the importance of engaging in productive dialogue. Overall, this paper deepens our understanding of how CCAs impact student dialogue and underscores the importance of a multidimensional perspective on CCA analyses to further enhance student collaboration.

AUTHOR CONTRIBUTIONS

Adelson de Araujo: Conceptualization; methodology; software; data curation; formal analysis; validation; investigation; visualization; writing – original draft; writing – review and editing; project administration. **Pantelis M. Papadopoulos:** Conceptualization; investigation; methodology; supervision; funding acquisition; writing – review and editing; validation; formal analysis; project administration; visualization. **Susan McKenney:** Conceptualization; methodology; writing – review and editing; funding acquisition; validation; supervision; investigation. **Ton de Jong:** Conceptualization; funding acquisition; writing – review and editing; supervision; resources; project administration; validation; methodology; investigation.

ACKNOWLEDGEMENTS

We like to express our sincere gratitude to teachers Eloize Pires, Juliano Almeida (from Colégio Antares Americana, Brazil), Gerbrand Wilbrink, Christien Tijhuis, and Martine Stevelink (from Canisius, the Netherlands) for their invaluable assistance in data collection, and to software engineers Shant Derounian and Jakob Sikken for integrating Clair with Go-Lab.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest related to this study. No external funding was received, and no outside organization influenced the design, execution, interpretation, or reporting of the research findings.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Adelson de Araujo https://orcid.org/0000-0002-4219-1955
Pantelis M. Papadopoulos https://orcid.org/0000-0002-1527-5483
Susan McKenney https://orcid.org/0000-0002-6028-4098
Ton de Jong https://orcid.org/0000-0003-0416-4304

REFERENCES

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, *37*(100489), 100489. https://doi.org/10.1016/J.EDUREV.2022.100489

Casalino, G., Castellano, G., & Zaza, G. (2022). Neuro-fuzzy Systems for Learning Analytics. *International Conference on Intelligent Systems Design and Applications, LNNS*, *418*, 1341–1350. https://doi.org/10.1007/978-3-030-96308-8_124

Chen, G., Zhang, J., Chan, C. K. K., Michaels, S., Resnick, L. B., & Huang, X. (2020). The link between student-perceived teacher talk and student enjoyment, anxiety and discursive engagement in the classroom. *British Educational Research Journal*, *46*(3), 631–652. https://doi.org/10.1002/BERJ.3600

Chi, M. T. H., & 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

Chinn, C. A., O'donnell, A. M., & Jinks, T. S. (2000). The structure of discourse in collaborative learning. *The Journal of Experimental Education*, *69*(1), 77–97. https://doi.org/10.1080/00220970009600650

Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2016). *21st century skills development through inquiry-based learning: From theory to practice* (1st ed.). Springer. https://doi.org/10.1007/978-981-10-2481-8

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*(1), 1–35. https://doi.org/10.3102/00346543064001001

de Araujo, A., Papadopoulos, P. M., McKenney, S., & de Jong, T. (2023a). Supporting Collaborative Online Science Education with a Transferable and Configurable Conversational Agent. *15th International Conference on Computer-Supported Collaborative Learning (CSCL)*.

de Araujo, A., Papadopoulos, P. M., McKenney, S., & de Jong, T. (2023b). Automated coding of student chats, a trans-topic and language approach. *Computers and Education: Artificial Intelligence*, *4*, 100123. https://doi.org/10.1016/J.CAEAI.2023.100123

de Jong, T. (2019). Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning*, *35*(2), 153–167. https://doi.org/10.1111/jcal.12337

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., Dziabenko, O., Koslowsky, J., Korventausta, M., Law, E., Pedaste, M., Tasiopoulou, E., Vidal, G., & Zacharia, Z. C. (2021). Understanding teacher design practices for digital inquiry–based science learning: The case of go-lab. *Educational Technology Research and Development*, *69*(2), 417–444. https://doi.org/10.1007/S11423-020-09904-Z

de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let's talk evidence—The case for combining inquiry-based and direct instruction. *Educational Research Review*, *39*, 100536. https://doi.org/10.1016/j.edurev.2023.100536

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gillies, R. M. (2019). Promoting academically productive student dialogue during collaborative learning. *International Journal of Educational Research*, *97*, 200–209. https://doi.org/10.1016/J.IJER.2017.07.014

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–student dialogue during classroom teaching: Does it really impact on student outcomes? *Journal of the Learning Sciences*, *28*(4–5), 462–512. https://doi.org/10.1080/10508406.2019.1573730

Hu, L., & Chen, G. (2023). A systematic review and meta-analysis of productive peer talk moves. *Journal of Behavioral Education*, 1–33. https://doi.org/10.1007/S10864-023-09513-9

Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, *4*(1), 21–34. https://doi.org/10.1109/TLT.2010.41

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Mercer, N. (2000). *Words and minds: How we use language to think together* (1st ed.). Routledge. https://doi.org/10.4324/9780203464984

Michaels, S., & O'Connor, C. (2012). Talk science primer.

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussions. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through talk and dialogue* (pp. 347–362). American Educational Research Association. https://doi.org/10.3102/978-0-935302-43-1_27

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*, *27*(4), 283–297. https://doi.org/10.1007/S11217-007-9071-1

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B. (2016). *Accountable talk: Classroom conversation that works*. University of Pittsburgh.

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. https://doi.org/10.1017/CBO9781139519526.026

Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education*, *192*, 104661. https://doi.org/10.1016/J.COMPEDU.2022.104661

Noroozi, O., Teasley, S. D., Biemans, H. J. A., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning*, *8*(2), 189–223. https://doi.org/10.1007/s11412-012-9162-z

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Rapanta, C., Miralda-Banda, A., Garcia-Milà, M., Vrikki, M., Macagno, F., & Evagorou, M. (2023). Detecting the factors affecting classroom dialogue quality. *Linguistics and Education*, *77*(101223), 101223. https://doi.org/10.1016/j.linged.2023.101223

Rojas-Drummond, S., Littleton, K., Hernández, F., & Zúñiga, M. (2010). Dialogical interactions among peers in collaborative writing contexts. In K. Littleton & C. Howe (Eds.), *Educational dialogues, understanding and promoting productive interaction* (pp. 128–148). Routledge.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 31–40. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education*, *188*, 104564. https://doi.org/10.1016/J.COMPEDU.2022.104564

Tegos, S., Demetriadis, S., Papadopoulos, P. M., & Weinberger, A. (2016). Conversational agents for academically productive talk: A comparison of directed and undirected agent interventions. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 417–440. https://doi.org/10.1007/s11412-016-9246-2

Valle Torre, M., Specht, M., & Oertel, C. (2023). *The sequence matters: A systematic literature review of using sequence analysis in Learning Analytics*. https://doi.org/10.48550/arXiv.2308.01218

Warner, J., Sexauer, J., Unnikrishnan, A., Castelão, G., Pontes, F. A., Uelwer, T., Batista, F., van den Broeck, W., Song, W., Badger, T. G., Pérez, R. A. M., Power, J. F., Mishra, H., Trullols, G. O., & Hörteborn, A. (2019). JDWarner/Scikit-Fuzzy: Scikit-Fuzzy Version 0.4.2. https://doi.org/10.5281/ZENODO.3541386

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers in Education*, *46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G. H., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2019). Multilingual Universal Sentence Encoder for Semantic Retrieval. ArXiv: 1907.04307. 87–94 https://doi.org/10.48550/arxiv.1907.04307

Zadeh, L. A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, *11*(1–3), 199–227. https://doi.org/10.1016/S0165-0114(83)80081-5

How to cite this article: A. de Araujo, Papadopoulos, P. M., McKenney, S., & T. de Jong (2024). A learning analytics-based collaborative conversational agent to foster productive dialogue in inquiry learning. *Journal of Computer Assisted Learning*, 1–15. https://doi.org/10.1111/jcal.13007