Temporal pathways to learning: how learning emerges in an open-ended collaborative activity

Jauwairia Nasir, Mortadha Abderrahim, Aditi Kothiyal, Pierre Dillenbourg

Angaben zur Veröffentlichung / Publication details:

Nasir, Jauwairia, Mortadha Abderrahim, Aditi Kothiyal, and Pierre Dillenbourg. 2022. "Temporal pathways to learning: how learning emerges in an open-ended collaborative activity." *Computers and Education: Artificial Intelligence* 3: 100093. https://doi.org/10.1016/j.caeai.2022.100093.

Computers and Education: Artificial Intelligence 3 (2022) 100093

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Temporal pathways to learning: How learning emerges in an open-ended collaborative activity.

Jauwairia Nasir *, Mortadha Abderrahim, Aditi Kothiyal, Pierre Dillenbourg

Computer-Human Interaction in Learning and Instruction (CHILI) Lab, Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland

ARTICLE INFO

Keywords:
collaborative problem solving
Collaborative learning processes
Multi-modal learning analytics
Hidden markov models
Time-series modelling
Productive engagement

ABSTRACT

The learning process depends on the nature of the learning environment, particularly in the case of open-ended learning environments, where the learning process is considered to be non-linear. In this paper, we report on the findings of employing a multimodal Hidden Markov Model (HMM) based methodology to investigate the temporal learning processes of two types of learners that have learning gains and a type that does not have learning gains in an open-ended collaborative learning activity. Considering log data, speech behavior, affective states and gaze patterns, we find that all learners start from a similar state of non-productivity, but once out of it they are unlikely to fall back into that state, especially in the case of the learners that have learning gains. Those who have learning gains shift between two problem solving strategies, each characterized by both exploratory and reflective actions, as well as demonstrate speech and gaze patterns associated with these strategies, that differ from those who don't have learning gains. Further, the teams that have learning gains also differ between themselves in the manner in which they employ the problem solving strategies over the interaction, as well as in the manner they express negative emotions while exhibiting a particular strategy. These outcomes contribute to understanding the multiple pathways of learning in an open-ended collaborative learning environment, and provide actionable insights for designing effective interventions.

1. Introduction

Learning does not occur in a single moment, but is rather a dynamic *process* that evolves over time (Kapur, 2011; Reimann, 2009). This process, especially in open-ended learning environments such as inquiry-based learning and problem-based learning environments, is non-linear (Brooks & Brooks, 1993; Chow et al., 2015; Schulte, 1996). Researchers have proposed that learning contexts are in fact complex systems where elements at different levels, such as cognitive, intrapersonal and interpersonal, interact and this results in the emergence of learning (Jacobson et al., 2016). Therefore, understanding the conditions for the emergence of learning in this complex system is important, as this will help identify those moments when an intervention could potentially be effective in improving learning. Within computer-supported collaborative learning (CSCL) research, there is now an emphasis to focus on how the CSCL process unfolds (Lämsä et al., 2021).

Further, learning is not a unimodal process and involves the interplay of cognition, emotions and actions. This is especially true in the case of collaborative learning which requires learners to sustain and regulate their cognition, emotions and actions in order to attain their goals (Järvelä et al., 2020). Previous research suggests that multimodal analysis can provide richer insights into the learning process compared to unimodal analyses (Nasir et al., 2021a; Blikstein & Worsley, 2016; Spikol et al., 2017). For instance, Olsen et al. (2020) found that combining eye gaze and audio modalities in a temporal analysis provides a more accurate prediction of collaborative learning than using single modalities alone. In Sinha (2021a), a multimodal learning analytic pipeline allows to not only infer the affective states that arise in a problem-solving followed by instruction (PS-I) activity, but also helps understand the temporal dynamics of such states and how they vary as the scaffolding strategies are manipulated.

While pre and post-tests help ascertain how much knowledge a learner has gained, they do not help understand *how* this knowledge was *gained* in a particular context, i.e., the temporal and multimodal aspects of the learning process. These aspects of the learning process have been previously studied using methods such as microgenetic analysis (Siegler & Crowley, 1991), interaction analysis (Jordan et al., 1995) and

* Corresponding author.
E-mail address: jauwairia.nasir@epfl.ch (J. Nasir).

https://doi.org/10.1016/j.caeai.2022.100093
Received 21 December 2021; Received in revised form 28 July 2022; Accepted 28 July 2022
Available online 24 August 2022

interactional ethnography (Castanheira et al., 2000) of learner discourse and actions, which track students conceptual development across an individual or collaborative learning activity. In an extensive review done by Chen et al. (2022) on the increasing use of Artificial Intelligence (AI) technologies in education, discourse analysis in CSCL was identified to be one of the most common trends and topics. However these qualitative methods can be time intensive. With technology-based learning contexts and multisensory data becoming increasingly widespread, researchers are making use of multiple sources of behavioral data such as interaction logs, audio, video, eye gaze and physiological data, along with machine learning methods, to understand the process of learning as a function of time (Engelmann & Bannert, 2021; Olsen et al., 2020). For example, in Lämsä et al. (2020), the authors make use of log data and lag sequential analysis to highlight the potential of temporal analysis to identify differences in the inquiry-based learning processes of scaffolded and non-scaffolded groups. Specifically, they discover three temporally distinct inquiry-based learning transition patterns among the three experimental groups that indicate different ways of using the scaffolds that could explain their learning. Further, in Csanadi et al. (2018), the authors show that their proposed methodology accounting for temporality, provides more insights than the traditional code-and-count strategies to characterize the socio-cognitive activities of learning in CSCL environments. Specifically, they found that 'evaluating evidence' was a core epistemic practice for dyads but not for individuals, suggesting that students collaborating argued in a more evidence-focused manner compared to individuals.

As demonstrated by the aforementioned studies, there is an increasing emphasis of AI in Education (AIEd). More specifically, authors in Chen et al. (2020), which is a systematic review of influential AIEd studies, found that "there was a lack of studies that both employ AI technologies and engage deeply with educational theories" and suggest to put more emphasis on understanding the relationship between learners answers (actions) and the underlying concepts. In this paper, our goal is then to develop such a temporal and multimodal understanding of the learning process in an open-ended collaborative activity seen in Fig. 1. Our work builds on our previous work Nasir et al. (2021c) which is grounded in theories of impasse-driven collaborative learning. *Therefore, we propose a Hidden Markov Model (HMM) based temporal analysis of multimodal behavioral data to unfold the differences and similarities between the collaborative learning processes of groups who learn and those who do not*. Our choice of using HMMs is motivated by the fact that HMMs allow us to model learning as a latent process based on our observations of student interaction with the learning activity, thus establishing a relationship between learner multimodal data and their collaborative learning.

In the upcoming section, we first review the literature regarding temporal and multimodal analysis methods for learning. Then in Section 3, we elaborate on the participants, the activity and the dataset used in this work, the experimental setup, as well as the adopted analysis methodology. This is followed by results, discussion, and conclusion in Section 4 and 5, respectively.

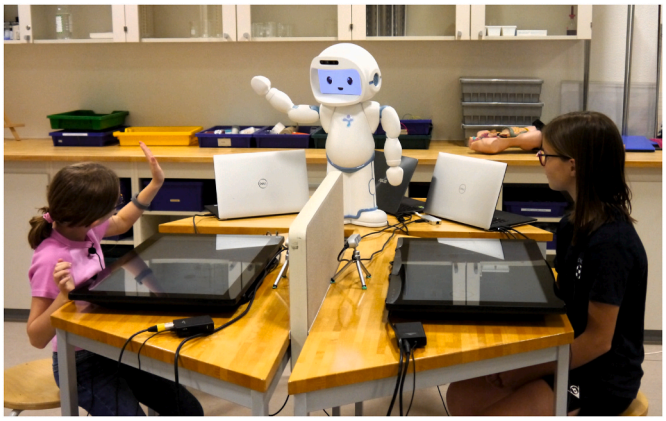


Fig. 1. A team interacting in an open-ended collaborative activity JUSThink.

2. Literature review

When embedded in a learning activity, intelligent agents must intervene at the right moment and in the right manner to enhance the learners' learning gains. Recent research suggests that student populations are diverse based on their levels of motivation, anxiety, autonomy, discipline and life experience (Lim, 2020). Therefore, we expect that these diverse populations learn in diverse ways and hence need to be supported differently from each other. Further, novel approaches to learning in the digital era such as citizen science initiatives (Ciasullo et al., 2022) and social media (Hosen et al., 2021) can result in different learning processes. As a result, in order to support learners better, we must have an ongoing comprehensive and deep understanding of the learners and learning situation. Temporal analysis of learners' data, either their performance or behaviors, can provide such an understanding.

2.1. Performance based systems

2.1.1. Knowledge Tracing

In Knowledge Tracing (KT) systems, temporal learner understanding is developed by estimating the learner's knowledge from their performance on past problems (Corbett & Anderson, 1994; Desmarais & Baker, 2012). **Bayesian Knowledge Tracing (BKT)** determines if and when the learning of a skill occurs during problem-solving steps (Desmarais & Baker, 2012). It assumes a two-state learning model where each skill is either in the learned or unlearned state. Assuming that each step of each problem calls for a single skill, the student can either succeed or fail the step, and the tutor updates its estimate of the learners knowledge on the skill accordingly (Corbett & Anderson, 1994; Desmarais & Baker, 2012). BKT has been applied both in the form of a *Hidden Markov Model* as well as in the form of a *Knowledge Tracing algorithm* (van de Sande, 2013). While these approaches have been applied successfully to model student knowledge in well-structured problem-solving, they fail at more complex open-ended learning activities (Wang et al., 2021). Hence, to increase the representational power and better model complex problem structures, Käser et al. (2017) suggest a **Dynamic Bayesian Network (DBN)** model that incorporates skill topologies. In this, different skills of a learning domain are considered within a single model capturing the dependencies between them. Incorporating skill hierarchies yields a significant improvement in predicting students' knowledge during complex problem solving, more accurately compared to the traditional KT models.

Further, **Deep Knowledge Tracing (DKT)** Piech et al. (2015), an application of recurrent neural networks, has been shown to be able to learn the latent structure in skill concepts without the need for explicit human coding of domain knowledge. For this reason, it demonstrates a drastic improvement on the well-known BKT models over several data sets. Nonetheless, similar to BKT, the DBN model as well as DKT assume that each problem-solving step or action maps to an underlying skill that could be either learned or unlearned, which is not necessarily the case in open-ended learning environments. Moreover, these approaches assume that an incorrect answer implies not learning or "slipping". However, it has been found that learners' actions that may seem to suggest failure vis-à-vis conventional standards of efficiency, accuracy, and performance quality may still lead to learning gains (Kapur & Kinzer, 2009). Thus, indicators other than in-task performance should be considered to model the learning process in open-ended learning activities. In Ramachandran, Huang, and Scassellati (2019), the authors suggest a link between motivation, actions, and the learning outcomes that underlies the learning process. They propose creating more effective tutoring interactions by finding observable behaviors that correspond to

motivational factors and employing a robot to respond to these behaviors. In Nasir et al., (2021a), the authors found that teams achieving higher learning gains in a robot-mediated human-human collaborative learning activity, may not necessarily perform well in the task. However, their speech, actions and emotions are distinctive as compared to the teams with lower learning gains. Thus, behavioral analysis could allow for better discrimination between high and low learners which will be the focus of our next sub-section.

2.2. Behavior based systems

2.2.1. Qualitative methods

When analyzing the learning process using learners' behaviors, both qualitative and quantitative approaches have been employed. Qualitative methods have been used to analyze, mainly, learners' gestures and speech to see how their learning is evolving. For instance, Jordan and McDaniel (2014) employ discourse analysis to describe the issues about which learners experienced uncertainty as they pursue collaborative learning projects that include a cognitive feeling of uncertainty. They identified how language was used in these particular social contexts to create and reflect meaning and structure. In Voutsina et al. (2019), authors used microgenetic task analysis to analyze the change in children's verbal reports when their overall solving approach appears to remain stable during a mathematical problem-solving task. They found that in fact the phases of stability are underlain by dynamic changes in the way the same strategy is communicated and conceptualized.

Although qualitative methods make it possible to contextualize and interpret the data based on human perception and analysis of the learning scenario, they sometimes overlook hidden factors that human observation cannot capture. Additionally, these methods are time and effort intensive, and as a result, do not scale up efficiently. With the development of sensors that capture data that is not perceivable by humans and the advancement in machine learning analysis techniques, there has been an increase in the deployment of quantitative approaches. Desmarais and Baker (2012) argue that as more and more learner data becomes available and methods for exploiting that data improve, there is potential for constant improvement of learner models. In this regard, researchers have attempted to gain an understanding of the learning process by considering multiple modalities and machine learning (ML) techniques as discussed below.

2.2.2. Quantitative methods

Perera et al. (2009) apply **sequential pattern mining (SPM)** on learners' log actions in a collaborative learning environment to extract sequences of frequent events. This analysis revealed interesting patterns, such as the presence of frequent task-focused communication, characterizing the teams ending up with positive and negative outcomes. Successful groups exhibit patterns suggestive of members giving frequent updates to the group while working on a task; such patterns are not present in the weaker groups. Kinnebrew et al. (2014) used SPM algorithms along with a hierarchical clustering algorithm to study the temporal evolution of the sequential patterns throughout the intervention, and compare the similarities and differences of their use between the experimental groups interacting with distinct versions of the software. The mined patterns allow for identifying and interpreting students' cognitive skills and learning behaviors. Besides, comparing these mined patterns with performance and context information, and tracking their temporal evolution better characterizes these behaviors as effective versus ineffective learning strategies. For instance, the importance of solution evaluation behaviors in complex learning tasks, is identified as one of the effective learning strategies.

Process Mining (PM) has also been applied to behavioral data to examine the learning process. This technique was adopted to discover the underlying problem solving or learning process model from the learning activity interaction sequence. Paans et al. (2019) employs a fuzzy miner algorithm, on sequences of encoded verbal utterances within dyads in a collaborative learning activity and find that repeated occurrences of social challenges during collaboration harm the learning outcomes. Here social challenges are defined as the failure to get along, a lack of joint attention, being highly critical, and so on. In fact, pairs, who repeatedly have disagreements, are more easily distracted, more easily go off-topic, have trouble getting back on topic again, and thus, are at risk for lower assignment quality.

Further, research suggests including more than one modality in the analysis because incorporating **multimodal techniques** would allow researchers to examine unscripted, constructionist, complex tasks in more holistic ways (Blikstein, 2013). Emerson et al. (2020) investigate this by analyzing log actions, facial expression of emotions, and eye gaze both separately and combined, and find that models utilizing multimodal data either perform equally well or outperform models utilizing unimodal data to predict learners' posttest performance and interest in a game-based learning environment. Olsen et al. (2020) further incorporate data temporality by using a **Long Short-Term Memory (LSTM)** model on log, gaze, audio, and dialog temporal data to predict teams' performance in a collaborative learning activity. The results indicate that combining various data streams from different time scales may be more beneficial than unimodal data. They also highlight the value of accounting for temporal aspects of the learning process as the temporal analysis of the gaze and audio measures provided accurate prediction of the normalized learning gain, while the averages and counts based analysis on the same features provided no information. Further, Giannakos et al. (2019) highlight how fused multimodal data, consisting of eye tracking, EEG, video, and wrist band data in addition to click stream data, can considerably reduce the prediction error for learning performance as compared to when only click streams are used in the design of learning technology. Lastly, in Yang et al. (2021, p. 2902), the authors have modelled the joint visual attention and with that the cognitive engagement of dyads using eye gazes and eye blinks data, and suggest that this multimodal temporal approach gives more and accurate insights into the collaborative problem solving engagement.

Another ML technique that has been used to temporally model the learning process with multimodal data is the **Hidden Markov Model (HMM)**. In Sharma et al. (2020), the authors use a combination of HMMs and the Viterbi algorithm to predict learners' effortful behaviors throughout the learning activity. They consider the effort categories as the hidden states and multimodal data-driven clusters as the observations. Results show that the suggested method outperforms the contemporary classification algorithms in classifying learners' behavioral patterns as effortful or effortless. Furthermore, this methodology highlights the exact moments when feedback is needed during the learning activity.

Literature suggests several data-driven multimodal ML approaches that could be used to analyze temporal data. Choosing a particular approach depends on the assumptions made about the measured data and the learning process underlying it, the nature of the data, the volume of available data, the purpose of the analysis, and the interpretability of the obtained models. The purpose of our analysis is to build a multimodal temporal model of the underlying process of learning as it happens in an open-ended collaborative learning activity. Sequence mining, sequential pattern analysis, and stochastic methods such as lag-sequential analysis, for instance, do not include the assumption of a latent learning process governing the sequence of observations (Bannert et al., 2014). Thus, we do not consider such methods for our temporal multimodal behavioral data analysis. Process mining, on the other hand, does account for latent processes; however, it is usually used to identify, confirm, or extend process models on sequential event data, which are sequences of discrete data, and thus, are different in nature from the data we investigate, which includes multivariate continuous features. Then, Recurrent Neural Networks (RNN), particularly LSTMs, have been broadly employed in order to analyze temporal multimodal behavioral data while complying with the assumption of a hidden process controlling the sequence of observations. Although promising (Spikol et al.,

2018), these neural networks lack the interpretability for multi-variable data regarding variable importance and variable-wise temporal importance due to their opaque hidden states (Guo et al., 2019). HMMs however offer more interpretability as the hidden states are well defined by their transition probabilities and emissions distributions. Therefore, they allow for a better understanding of the latent learning process during the learning activity.

Therefore, in this paper, we adopt the approach of building a Hidden Markov Model of the learning process, trained on learners' multimodal behavioral data. Our goal is to examine how these behaviors evolve throughout the activity and lead to learning gains during an open-ended collaborative learning activity. Broadly, our research question is, "*How do the learning behaviors of different types of learners evolve across an open-ended collaborative learning activity?*"

3. Methods

3.1. Participants

We make use of data from a previous study conducted with a robot-mediated open-ended collaborative learning activity called JUSThink (Nasir et al., 2020). The study[1] was conducted in two international schools in Switzerland over two weeks. A total of 96 learners aged 9–12 years old participated in the study. The participants were organized in teams of two, resulting in a total of 48 teams. However, to ensure data completeness and homogeneity, only data from 32 teams were used for this analysis. Specifically, all teams with incomplete or lost data in terms of log actions, audio or video data, pre/post tests were removed, leaving us with 34 teams (from which we generated our dataset elaborated in a section 3.4). Further, we removed two more teams that were outliers in terms of their behaviors (based on data driven behavior profiles that were generated in an earlier work as will be explained in section 3.5).

3.2. Activity

JUSThink aims to:

- improve children's computational skills by providing intuitive knowledge about minimum-spanning-tree problems
- promote collaboration among the team via its scripted design.

The learning task introduces the minimum-spanning tree problem through a gold mining scenario based on a map of Switzerland, where mountains represent gold mines and are nodes that should be connected by railway tracks, representing the edges, that each have a cost to build. The robot, playing the role of the CEO of a gold-mining company, restates the problem by asking learners to help it collect the gold by connecting the mines with railway tracks. The participants must collaboratively construct the solution by connecting the mines while spending as little money as possible on building the railway tracks. Our motivation for choosing the minimum spanning tree problem and computational thinking skills as the domain for this collaborative activity is based on the recent push towards introducing CT skills in early education (Menon et al., 2019) as well as the idea that robots could be one possible effective tool for advancing these skills (Chalmers, 2018). Further, in the process of organizing this study, we received feedback from various teachers that such an activity can be complementary to the curriculum on optimization problems taught to the targeted age range; hence, this motivated our choice for the age range of students as mentioned in Section 3.1.

We chose to have an open-ended collaborative activity where learners collaborate to solve an open-ended problem without receiving direct guidance, and this is inspired by the inherent characteristic of such problem-solving followed by instruction (PS-I) activities that encourage the awareness of knowledge gaps, stimulate knowledge construction processes and lead to increased learning gains (Loibl et al., 2017; Sinha & Kapur, 2021). Additionally, it is known that collaborative activities need to be scripted for better collaboration and learning (Kollar et al., 2006; Vogel et al., 2017). Therefore, we designed a script based on partial information, role switching and complementarity. Concretely we implemented it by having two different views in the task: a figurative view and an abstract view, which provide complementary functionality as each gives only partial information to the user. On the one hand, the nodes and edges of the graph are shown as mountains and railway tracks in the figurative view. In this view, one can build and erase tracks. On the other hand, the abstract view has nodes and edges as circles and solid lines respectively, and deleted railway tracks are shown with dashed lines along with their cost so that one can view the cost of every track ever added (costs are revealed only when a track is first added). The learners can also access previous solutions and their costs and bring back a previous solution. Given the nature of the problem and the number of views, collaboration in twos was optimal for this scenario. Hence we had teams of two, with these two views being swapped between participants every two moves, enabling both team members to experience the thought process that comes with the view. Given this collaborative script, team members need to communicate in order to use the information in both the views, make decisions and build the solution. Furthermore, they need to agree on a solution spanning the whole graph, as they both need to press the submit button for it to be submitted to the robot for evaluation. The robot intervenes intermittently during the learning task to provide feedback on the progress, give hints, and lend support through minimal verbal and non-verbal behaviors. More on the task can be found in (Nasir et al., 2020).

Teams of two children each took part in the activity that lasted approximately 50 min. First, the robot welcomes the children and explains the goal of the task. Participants then take an individual pre-test. Following the pre-test, the robot introduces the two game views and their functionalities. The learning task then begins and lasts around 25 min, after which, children complete an individual post-test and a self-assessment questionnaire. Finally, the robot greets them goodbye. The robot thus mediates and automates the entire activity by giving instructions and by moving the activity from one stage to the next as required. It also provides some motivational feedback along the way.

The pre and post-tests consist of questions with a context other than the learning task scenario and are based on variants of the graphics in the muddy city problem.[2]

3.3. Experimental setup

As seen in Fig. 1, the two team members sit across from each other with a touch screen placed horizontally in front of each one. They are separated by a barrier so as to be able to see each other but not each other's screen. The humanoid robot (QTrobot) is placed on the side visible to both children. Data was collected throughout the activity using one environment camera to capture the whole interaction scene, two RGB-D front cameras, one for each participant to capture the face up-close, and two lavalier microphones to capture audio data. Two computers, connected to the screens and the robot, manage the activity and the synchronous recording of the sensors.

Each team member interacts with an instance of the JUSThink application. A separate robot application manages the robot. All of the applications communicate via Robot Operating System (ROS). Participants' and robot's actions are recorded using Rosbags.

[1] This study received the approval of the university's ethics committee with reference number HREC No.: 051-2019.

[2] https://classic.csunplugged.org/activities/minimal-spanning-trees/.

3.4. Dataset

We make use of our open-source dataset PE-HRI_Temporal (Nasir et al., 2021b) generated from the data collected in the study mentioned in section 3.1. In the data set, for each team, the interaction of around 20–25 min is organized in windows of 10 s; hence, we have a total of 5048 windows of 10 s each. We report team level log actions, speech behavior, affective states, and gaze patterns for each window. More specifically, within each window, 26 features are reported in two formats; hence, giving a total of 52 values. We make use of the *non-incremental* format of the 26 features which means we look at the value of a feature in that particular time window without carrying any information from previous time windows. For more details, please see Nasir et al. (2021b). The 26 features are listed in Tables 1–3. The rationale for using these features to analyze learning are explained in our previous publication (Nasir et al., 2021a).

In addition to the features mentioned in Table 1, each window also includes a *normalized_time* feature which refers to the time when this window occurred with respect to the total duration of the task for a particular team. The dataset also consists of team level learning and performance metrics, where performance is measured based on the cost of a current solution relative to the optimal solution, while learning gains (absolute, relative or joint-absolute) are calculated by looking at the difference between the students scores on their post-tests and pre-tests. More detailed definitions are provided at Nasir et al. (2021b). Please note again that this dataset provides data for 34 teams, but for our current analysis we make use of data from 32 teams, as mentioned previously, giving us 4676 windows. Lastly, considering learning analytics and/or educational human-robot interaction studies with a robot, similar or even lower sample sizes are the norm (Belpaeme et al., 2018; Gordon et al., 2016; Ramachandran, Sebo, & Scassellati, 2019), as is the case with the type of analysis that we do in this work (for example, see Sharma et al. (2020).

3.5. Analysis methodology

Since the methodology of this paper builds on the outcomes of our previous work (Nasir et al., 2021c), we briefly describe it here. In the earlier work, we generated behavioral profiles based on the same features described above in section 3.4, but aggregated across the entire activity. We found differences in the behaviors between those who learn, i.e., *gainers* and those who do not end up learning, i.e., *non-gainers*. Further, we also observed behavioral differences in the two types of gainers (Nasir et al., 2021c). We saw that while *speech behavior* was a discriminatory factor between gainers and non-gainers, it was actually the interplay between problem solving strategies and emotional expressivity that distinguished the different ways in which gainers learned.

Table 1
Log features from our PE-HRI-Temporal dataset.

Log Features	
Feature Name	Description
T_add	The number of times a team added an edge on the map in that window
T_remove	The number of times a team removed an edge from the map in that window
T_ratio_add_rem	The ratio of addition of edges over deletion of edges by a team in that window
T_action	The total number of actions taken by a team (add, delete, submit, presses on the screen) in that window
Redundant_exist	The number of times the team had redundant edges in their map in that window
T_hist	The number of times a team opened the sub-window with history of their previous solutions in that window
T1_T1_add	The number of times either of the two members in the team followed the pattern consecutively: I delete an edge, I add it back in that window
T1_T1_rem	The number of times either of the two members in the team followed the pattern consecutively: I add an edge, I then delete it in that window
T1_T2_add	The number of times the members of the team followed the pattern consecutively: I delete an edge, you add it back in that window
T1_T2_rem	The number of times the members of the team followed the pattern consecutively: I add an edge, you then delete it in that window
T_help	The number of times a team opened the instructions manual in that window

Table 2
Video based features from our PE-HRI-Temporal dataset.

Video Features: Affective states and Gaze	
Feature Name	**Description**
Positive_Valence	The average value of positive valence for the team in that window
Negative_Valence	The average value of negative valence for the team in that window
Difference_in_Valence	The difference of the average value of positive and negative valence for the team in that window
Arousal	The average value of arousal for the team in that window
Gaze_at_Partner	The average of the two team member's gaze when looking at their partner in that window where each individual member's gaze is calculated as a percentage of time in that window.
Gaze_at_Robot	The average of the two team member's gaze when looking at the robot in that window where each individual member's gaze is calculated as a percentage of time in that window.
Gaze_other	The average of the two team member's gaze when looking in the direction opposite to the robot in that window where each individual member's gaze is calculated as a percentage of time in that window.
Gaze_at_Screen_Left	The average of the two team member's gaze when looking at the left side of the screen in that window where each individual member's gaze is calculated as a percentage of time in that window.
Gaze_at_Screen_Right	The average of the two team member's gaze when looking at the right side of the screen in that window where each individual member's gaze is calculated as a percentage of time in that window.
Gaze Ratio of Screen_Right and Screen_Left	The average ratio of a team member looking at the right side of the screen over the left side in that window

Table 3
Audio based features from our PE-HRI-Temporal dataset.

Audio Features: Speech	
Feature Name	**Description**
Speech_Activity	The average of the two team member's speech activity in that window where each individual member's speech activity is calculated as a percentage of time that they are speaking in that window.
Silence	The average of the two team member's silence in that window where each individual member's silence is calculated as a percentage of time in that window.
Short_Pauses	The average of the two team member's short pauses over their speech activity in that window. Each individual member's short pause refers to a brief pause of 0.15 s and is calculated as a percentage of time in that window.
Long_Pauses	The average of the two team members long pauses over their speech activity in that window. Each individual member's long pause refers to a pause of 1.5 s and is calculated as a percentage of time in that window.
Speech_Overlap	The average percentage of time the speech of the team members overlaps in that window.
Overlap_to_Speech_Ratio	The ratio of the speech overlap over the speech activity of the team in that window.

Based on that, we identified the two types of gainers as *Expressive Explorers* and *Calm Tinkerers*, and the non-gainers as *Silent Wanderers*. In this paper, we retain the same terminology. While the aforementioned behavioral profiles highlight the aggregate differences between all types of learners, in order to identify the differences between the *learning process* of those who learn and those who do not, we employ HMMs to generate multi-modal *temporal* behavioral profiles for each type of learners. This enables us to understand how the multimodal behaviors of each type of learners evolve throughout the interaction.

An HMM is a doubly stochastic model with an underlying stochastic process that is not observable, but can only be observed through another set of stochastic processes that produce the sequence of observed symbols. It is specified by a set of N states, an initial probability distribution, a transition probability matrix, and a sequence of emission probabilities. Additionally, HMMs require three assumptions: firstly, that the next state is dependent only on the current state, secondly, that the state transition probabilities are independent of the time of transition and finally, that the current observations are statistically independent of the previous outputs. In our case, our data is grouped into independent 10 s windows, with each window containing behaviors occurring in those 10 s alone, and thus assumption 3 holds. Further, each hidden state of the HMM manifests a set of significantly different behaviors by which the state is characterized; this set of behaviors together signify a particular *approach to learning*. Hence, the next state or the approach to learning taken next by a pair of learners depends only on the current state (assumption 1) and the probability of transitioning to a different approach to learning is independent of when in the activity it occurs (assumption 2). Thus all the assumptions required to do an HMM analysis are valid for our data and learning context; hence, allowing us to proceed with HMM modeling. Our analysis consists of four main steps:

3.5.1. Step1: ***Preprocessing***

As our features come from different kinds of behavioral modalities, they are on different scales. So we begin by applying a min-max scaler to normalize our data.

3.5.2. Step2: ***Behaviors Clustering***

In order to have a starting point for the number of states of the HMM, we perform a clustering of the temporal behavioral features to identify significantly different behavioral clusters. We then assume that these clusters are emitted by distinct hidden states, and so the number of states is the same as the number of behavioral clusters. For clustering, a Principal Component Analysis (PCA) is conducted to compute the principal components, the first components are kept based on the elbow method on the proportion of variance explained. The Principal Components are then clustered using the K-Means algorithm. The number of clusters is optimized based on the elbow method on inertia and the silhouette score. In order to confirm that the obtained clusters are actually different in terms of multimodal behaviors, we perform a Kruskal-Wallis test on the clusters' behavioral features. This test further serves as a means to identify behaviors that significantly distinguish a cluster from the other. This step is summarized in Fig. 2.

3.5.3. Step3: ***the HMM***

Since our temporal behavioral features are multivariate and most of them have continuous values, our emission probability distribution should be continuous multivariate. Thus, for this step, we use the GMMHMM model provided by the hmmlearn library,[3] as it accounts for the aforementioned condition by representing the emission distribution as a mixture of multiple Gaussian densities.

We set the number of hidden states to the number of clusters found in the previous step. The HMM is then trained using the Expectation-Maximization algorithm on the set of the teams' sequences. Each sequence consists of all the observations of a team sorted in increasing order of time, where an observation consists of the normalized multimodal behavioral features and time at a given time window. We then apply the Viterbi algorithm on these sequences to recognize at which hidden state each observation is emitted. As a result, for each hidden state, we can construct the set of observations emitted by that state. Finally, we perform a Kruskal-Wallis test on each feature between each pair of these sets with the significance threshold set to 0.01. For each of the significantly different features between a pair of sets, we further compare the mean values across the sets and label the mean value of each set with one of the labels {Highest, High, Medium, Low, Lowest} based on a generated score in the following manner:

For a significantly different feature x, we first define:

$$min(x) = minimum\ of\ mean\ values\ of\ x\ across\ all\ sets$$

$$max(x) = maximum\ of\ mean\ values\ of\ x\ across\ all\ sets$$

Then, for a set i, we generate a score for the feature x as:

$$score(x, i) = \frac{(mean\ of\ x\ in\ i - min(x))}{(max(x) - min(x))}$$

Lastly, the feature x in i is labeled with:

- 'Highest', if $score(x, i) = 1$.
- 'High', if $2/3 \leq score(x, i) < 1$.
- 'Medium', if $1/3 \leq score(x, i) < 2/3$.
- 'Low', if $0 < score(x, i) < 1/3$.
- 'Lowest', if $score(x, i) = 0$.

The significantly different features and their labels for a set i represent the manifestation of the hidden state corresponding to the set i and we subsequently use these labeled features to represent the state. This enables us to interpret the progression of the hidden learning states in terms of the values of the significantly differing observed behaviors. Fig. 3 outlines the processes employed to train and interpret the model.

In conclusion, in this step, the HMM is trained in order to learn the hidden states that emit the observed multimodal behavioral features, and the significantly different features that characterize each state are identified. Interpreting these results allows for building the learning profiles that dyads go through during the activity. Furthermore, the model allows for learning the initial probability distribution as well as the probabilities to transition from one state to the other, which allows for building the temporal profile.

This entire pipeline, as summarized in Fig. 4, is adopted to identify the temporal profiles for each type of learners separately.[4] Its implementation is publicly made available in a Github repository.

4. Results

This section presents the results of the analysis methodology applied to the temporal multi-modal datasets of the *Expressive Explorers*, the *Calm Tinkerers*, and the *Silent Wanderers*. The clustering analysis, as discussed in the previous section, applied for the *Expressive Explorers*, the *Calm Tinkerers*, and the *Silent Wanderers* suggests the following number of components [PCs = 4, PCs = 4, PCs = 5 respectively] and the following number of clusters [K = 2, K = 3, K = 3 respectively], based on the elbow method on inertia and the silhouette scores. These are considered as a starting point for the number of hidden states, and we further train Hidden Markov models with K+1 states to identify whether other non trivial states exist or not, that eventually suggests that we have three hidden states for each of these groups. Hence, we define the

[3] hmmlearn is a set of algorithms for unsupervised learning and inference of Hidden Markov Models, https://hmmlearn.readthedocs.io/.

[4] Github repository': https://github.com/chili-epfl/justhink-HMM.

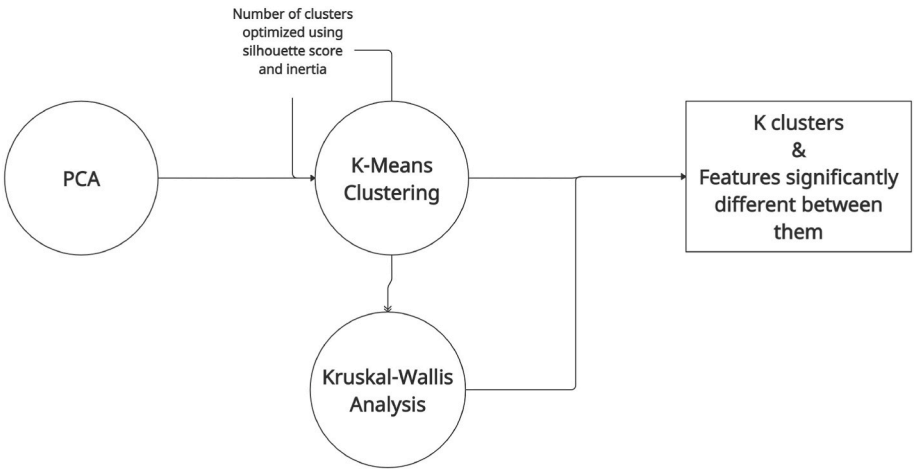


Fig. 2. Behaviors Clustering step.

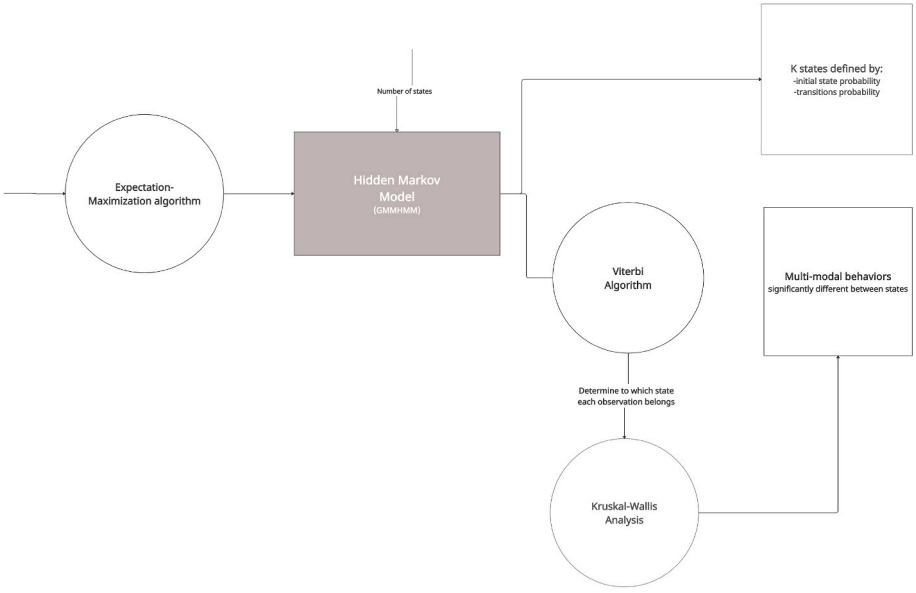


Fig. 3. The HMM step.

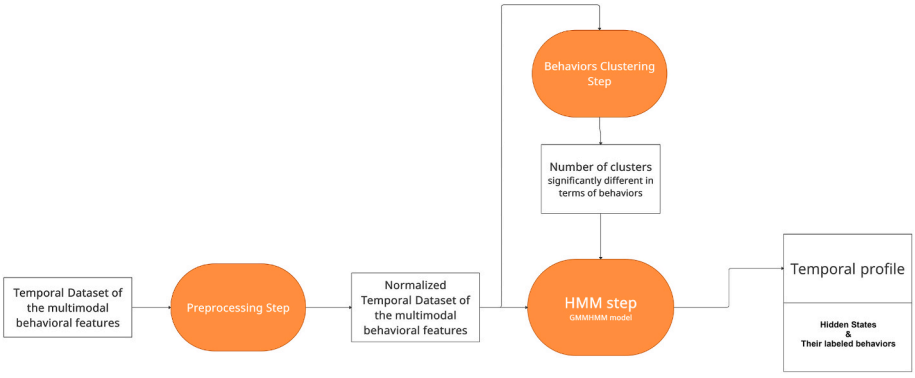


Fig. 4. The analysis methodology.

following naming convention for the hidden states in each of the groups' models:

- ***InitialState***: the state with the highest initial probability.
- ***MoreProbableState***: the state with the highest transition probability from the initial state.
- ***LessProbableState***: the state with the lowest transition probability from the initial state.

We further define the following conventions for the state diagrams:

- The size of a state in the state diagrams is representative of its initial state probability. That is, the bigger the circle representing the state, the bigger its initial probability is.
- The size of the font of the transition probabilities in the state diagrams is illustrative of its magnitude. Explicitly, higher transition probabilities have bigger font sizes.

For each of the three groups, their HMM model, trained on sequences of observations of the respective group and the number of states set to three, is represented by the state diagrams in Figs. 5–7, respectively. For all groups, the probabilities suggest that once in *InitialState*, staying in that state has the highest probability compared to other possible transitions. However, once out of this state, going back to the *InitialState* from the *LessProbableState* and *MoreProbableState* generally has lower transition probabilities. The probabilities are especially low in the case of *Expressive Explorers* from both of the other states, and for both *Calm Tinkerers* and *Silent Wanderers* from the *LessProbableState*. On the other hand, the *Silent Wanderers* can still transition from *MoreProbableState* to *InitialState* with a non-trivial probability of *0.305* which is higher than the probability of going to *LessProbableState* from *MoreProbableState*. Similarly, the *Calm Tinkerers* also have a relatively higher transition probability to go back to the *InitialState* from their *MoreProbableState*; however, they still have a higher probability to transition to their *LessProbableState* from this state. Furthermore, the findings from the Kruskal-Wallis analysis comparing the values of the multi-modal behavioral features between each pair of states, for each group of learners, is shown in tables next to the respective HMM models. The tables include the features which represent the manifestation of the hidden states. Note that the features that do not differ significantly between the states are not shown in these tables. This does not mean the absence of that feature in a state, rather that the feature does not differ significantly between states, i.e., the value of that feature does not oscillate between states significantly. We discuss further on these results in the upcoming section.

5. Discussion

5.1. Temporal multimodal behavioral profiles

In this section, we describe the higher level understanding that the temporal analysis, based on the HMMs identified in the previous section, provides us of how the multi-modal behaviors of each group of learners evolve during the collaborative learning activity and what this says about their learning process. Based on the findings in Section 4, we observe two kinds of problem solving (PS) strategies namely:

- Global PS Strategy: This strategy includes global level exploration and/or reflection characterized by addition actions and looking at past solutions (history).
- Local PS Strategy: This strategy includes local level exploration and/or reflection characterized by deletion actions and addition followed by deletion actions or vice versa.

Previously, in the results section, we name our states on the basis of initial probability (*InitialState*) or transition probabilities from the initial state (*LessProbableState*, *MoreProbableState*). In this section, we try to understand the nature of the states and consequently, we name them based on their:

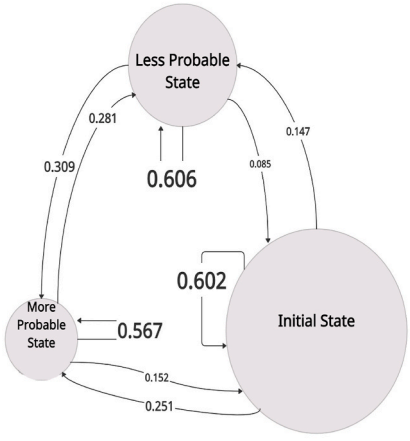
1. Productivity
2. Problem solving strategy

With respect to 1, in our previous work (Nasir et al., 2021c), we found that the quantity and quality of speech was able to discriminate between productive and non-productive teams in terms of learning. Additionally, we found that when the behaviors were averaged across the entire interaction for each team, there were two problem solving strategies (Global PS Strategy and Local PS Strategy) that emerged and overall, one group of gainers displayed only one strategy, while the other group of gainers displayed the other. However, the temporal profiles of each group of learners help elaborate these findings further.

Please note that in the upcoming figures of the profiles, the strength of the transition probabilities is represented by the strength of the arrows and the unproductive, semi-productive and productive states and transitions are represented by the different colors as described in the legend of the figures.

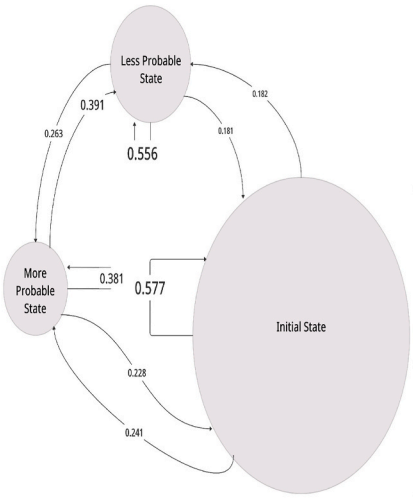
5.1.1. Expressive Explorers

The temporal profile for *Expressive Explorers* is shown in Fig. 8 from which we see that these learners start, with the highest probability, at a state characterized by more technical help-seeking, fewer actions with the learning activity, and high silence. For these reasons, it appears to be



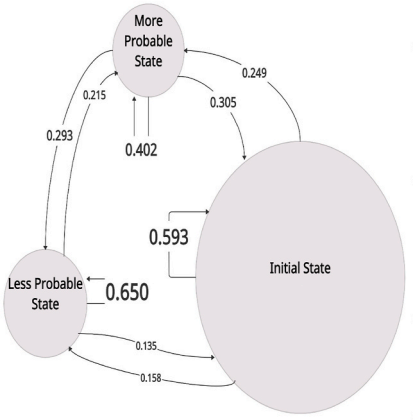
Featuraes	States		
	Initial State	More Probable State	Less Probable State
Log Features:			
T_add	Lowest	Highest	Low
T_remove	Lowest	Low	Highest
T1_ratio_add_rem	Lowest	Highest	Low
T_action	Lowest	Highest	Low
Redundant_exist	Low	Highest	Lowest
T_hist	Highest	High	Lowest
T1_T1_rem	Low	Lowest	Highest
T1_T2_rem	Lowest	Low	Highest
T_help	Highest	Lowest	Low
Video Features:			
Negative_Valence	Lowest	Medium	Highest
Arousal	Lowest	Medium	Highest
Gaze_at_Partner	Low	Lowest	Highest
Gaze_at_Robot	High	Lowest	Highest
Gaze_at_Screen_Left	Highest	High	Lowest
Gaze_at_Screen_Right	Highest	High	Lowest
Audio Features:			
Speech_Activity	Lowest	High	Highest
Silence	Highest	Low	Lowest
Short_Pauses	Highest	High	Lowest
Long_Pauses	Low	Highest	Lowest
Speech_Overlap	Lowest	High	Highest
Overlap_to_Speech_Ratio	Lowest	Medium	Highest
Normalized_Time	Lowest	High	Highest

Fig. 5. HMM State diagram for the Expressive Explorers.



Featuraes	States		
	Initial State	More Probable State	Less Probable State
Log Features:			
T_add	Lowest	Highest	Low
T_remove	High	Lowest	Highest
T_ratio_add_rem	Lowest	Highest	Low
T_action	Lowest	Highest	Low
Redundant_exist	Lowest	Highest	Low
T1_T1_rem	Low	Lowest	Highest
T1_T2_rem	Low	Lowest	Highest
T_help	High	Lowest	Highest
Video Features:			
Positive_Valence	Lowest	Highest	Medium
Negative_Valence	Lowest	Highest	High
Difference_in_Valence	Lowest	Highest	Low
Arousal	Lowest	Highest	High
Gaze_at_Partner	Highest	Lowest	Medium
Gaze_at_Robot	Medium	Lowest	Highest
Gaze_at_Screen_Left	Lowest	High	Highest
Gaze_at_Screen_Right	High	Highest	Lowest
Audio Features:			
Speech_Activity	Lowest	High	Highest
Silence	Highest	Low	Lowest
Short_Pauses	Highest	High	Lowest
Long_Pauses	Highest	Medium	Lowest
Speech_Overlap	Lowest	High	Highest
Overlap_to_Speech_Ratio	Lowest	High	Highest
Normalized_Time	Lowest	Highest	High

Fig. 6. HMM State diagram for the Calm Tinkerers.



Featuraes	States		
	Initial State	More Probable State	Less Probable State
Log Features:			
T_add	Low	Highest	Lowest
T_remove	High	Lowest	Highest
T_ratio_add_rem	Low	Highest	Lowest
T_action	Low	Highest	Lowest
Redundant_exist	Medium	Highest	Lowest
T_help	Highest	Low	Lowest
Video Features:			
Positive_Valence	Lowest	Highest	High
Difference_in_Valence	Low	Lowest	Highest
Arousal	Lowest	Highest	Medium
Gaze_at_Partner	Lowest	Highest	Medium
Gaze_at_Screen_Left	Highest	Medium	Lowest
Gaze_at_Screen_Right	Lowest	Highest	High
Audio Features:			
Speech_Activity	Lowest	Medium	Highest
Silence	Highest	Medium	Lowest
Speech_Overlap	Lowest	Low	Highest
Overlap_to_Speech_Ratio	Lowest	Low	Highest
Normalized_Time	Lowest	Medium	Highest

Fig. 7. HMM State diagram for the Silent Wanderers.

a state of non-productivity. As opposed to the averages and frequency analysis in Nasir et al. (2021c), which suggests that *Expressive Explorers* learned by following a more global problem solving strategy, this temporal analysis indicates that once they go out of the non-productive state, they employ both of the problem solving strategies: in the more probable state they follow a global problem solving strategy of adding edges and looking more at their previous solutions, and a less probable state where they follow a local problem solving strategy consisting of more removals in general, and removing each other's last added edges in particular. What is interesting is that the latter state is more likely to occur at later times in the activity than the global problem solving state, suggesting that these students begin with a more global problem solving approach and move on to a more local strategy of making quick changes. This transition is also characterized by increasing negative emotions, such as frustration, that is perhaps brought on by the awareness of reaching the end of the activity and the allotted time. In the states of non-productivity (while trying to understand the activity) and global problem solving (while adding edges), the learners gaze at the screen is high, while in the state of local problem solving while removing edges, and in particular each others' edges, the learners gaze at their partners is highest. However, both of the problem solving states are characterized by high speech and speech overlap which signifies good collaboration (Viswanathan & Vanlehn, 2018). Once *Expressive Explorers* reach a productive state, it is highly unlikely to get back to the non-productive one.

5.1.2. Calm Tinkerers

Calm Tinkerers as shown in Fig. 9 start, with the highest probability, at a state characterized by high technical help-seeking, fewer actions, and high silence. Due to these behaviors, it seems to be a state of non-productivity. Similar to *Expressive Explorers*, the temporal analysis done in this paper gives a richer insight into these learners behaviors. Contrary to the aggregate analysis which suggested that these learners adopt a local problem solving strategy, this analysis suggests that these type of gainers too go through two states of productivity: a less probable state of local problem solving and a more probable state of global problem solving. In the state of local problem solving, *Calm Tinkerers* do most removal actions, particularly removing each other's last added edges, show lesser negative emotions, and their speech is at its highest. In the state of global problem solving, these learners do more addition actions, are more frustrated and their speech decreases but is still relatively high. Contrary to *Expressive Explorers*, we find that in *Calm*

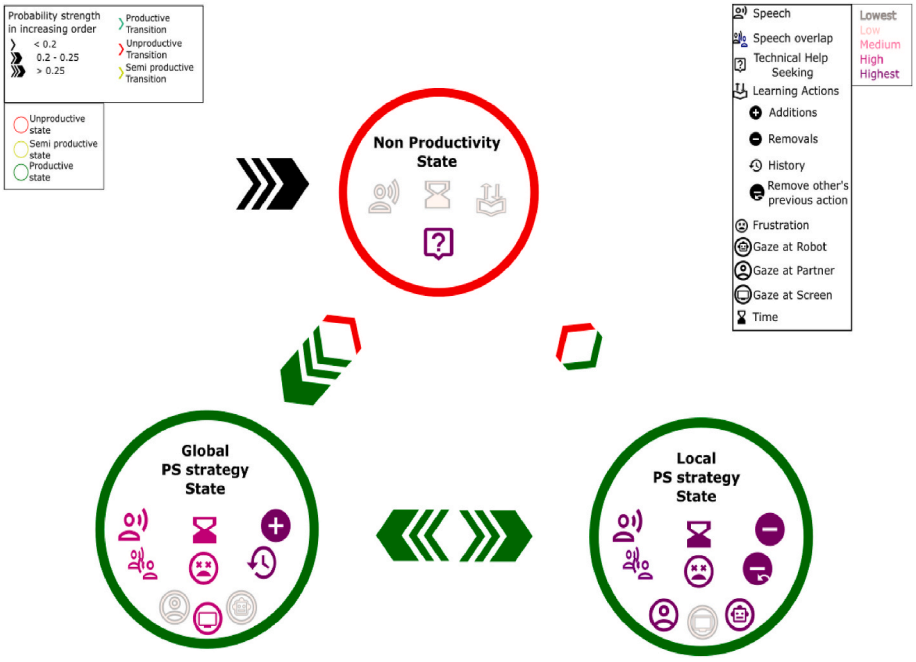


Fig. 8. Temporal profile for Expressive Explorers.

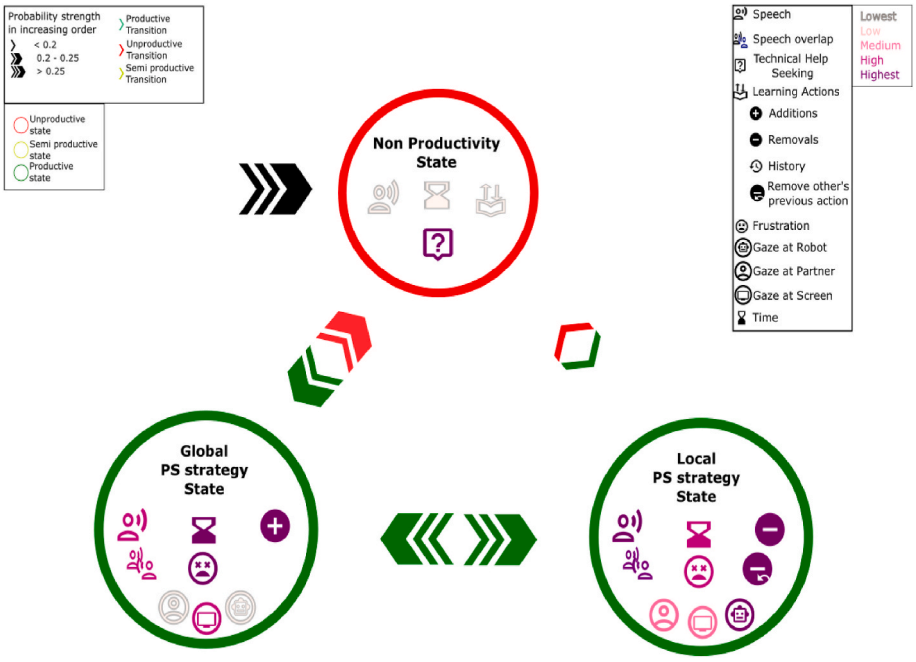


Fig. 9. Temporal profile for Calm Tinkerers.

Tinkerers the state of local problem solving is more likely to occur earlier in the activity than the state of global problem solving, suggesting that these learners begin with a local problem solving approach. However, similar to the *Expressive Explorers*, these learners change in problem solving strategies is also accompanied with an increase in negative emotions (see Fig. 9).

In the state of non-productivity while trying to understand the activity, the *Calm Tinkerers* gaze at their partner as well as the right side of the screen is high, in the state of global problem solving while adding edges the learners gaze on both sides of the screen is high and in the state of local problem solving while removing edges, including each others' edges, the learners gaze at the robot and the left side of the screen is highest. We must note that the only difference between the left and the right sides of the screen is that if a previous solution is opened, it is displayed on the right side; whereas, the information on the total number of nodes and the number of edges currently present on the map is on the left side. Similar to *Expressive Explorers* both the productive states are characterized by high speech signifying good collaboration in both states (Viswanathan & Vanlehn, 2018). Further, similar to *Expressive Explorers*, the speech in the local PS state is highest and this is likely because this state involves the highest removal of each others' edges which requires discussion and agreement among both partners, thus increasing the speech activity. Lastly, different from *Expressive Explorers*, these learners still have a medium probability to fall back to

the unproductive state from the state of global problem solving strategy.

5.1.3. Silent Wanderers

Similar to the two gainer groups, the *Silent Wanderers* (shown in Fig. 10) start with the highest probability at a non-productive state characterized by more technical help-seeking, high silence, and low actions with the learning activity. They go through a more probable state, occurring in the middle of the activity (suggested by medium normalized time), where they adopt a global problem solving strategy in which their speech increases and they do more addition actions. However there is no change in their reflective actions in this state, either in terms of looking at their previous solutions or removing their own or their partners added edges. Even from this state of productivity, they can still fall back to the state of non-productivity with a high transition probability. In the less probable state, which is more likely to occur towards the end of the activity and is characterized by a more local problem solving strategy, non-gainers do more removals and few additions. We may infer that this is a more reflective phase although their reflection, unlike the gainers, does not include a significant increase in the use of the solution history or each other's last actions. However, this state is characterized by their highest speech.

In terms of gaze, in the non-productive state while trying to understand the activity these learners gaze at the left side of the screen is highest and this could be because the information on the number of nodes and the number of edges currently present on the map is located on the left. In the more probable state of doing additions, their gaze at their partner and the right side of the screen is highest, where the history is also located and it could be that learners were accessing their past solutions. Finally, in the less probable state of removing edges, their gaze at the right side of the screen is high, which could again indicate learners accessing their history. Interestingly, we find no difference in the learners frustration between the three states, indicating that their negative emotions were relatively stable regardless of whatever they were doing in the activity. Thus our analysis reveals that non-gainers go through a "slower" learning pathway characterized by an intermediate semi-productive state where actions on the activity and speech increases, but reflection is generally unchanged. While they do reach a productive state of reflective problem solving and higher amount of discourse, it is reached late in the activity. However, this suggests that given time even the non-gainers could achieve higher learning gains since once they reach this productive state, similar to gainers, the probability of going back to the non-productive states is low. We hypothesize that the lack of reflection in the intermediate state could be the reason why non-gainers do not have higher learning gains as it is known that reflection plays a crucial role in learning from problem solving (Do-lenh, 2012; Hmelo-Silver, 2004).

Together our findings suggest that not only are there multiple behavioral profiles of learning (Nasir et al., 2021c), there are multiple behavioral pathways for learning, and learners who have learning gains do not adopt a single problem solving strategy, global or local, but indeed a combination of both. Further, they modify strategies based on the status of the problem solving and feedback obtained from the environment. Our findings also suggest an interplay between PS strategies and other behaviors which we explore in-depth in the next section.

5.2. Interplay between PS strategies and other behaviors

Now that the temporal learning profiles have been explained for each group, we would like to focus on how speech, affect and gaze evolve for each of these groups and interplay with the global vs the local problem solving strategies i.e., while performing addition actions predominantly or when removal actions are more frequent, respectively. This *interplay* between the *problem solving strategies* and behaviors of *speech*, *gaze*, and *affect* is shown in Table 4, which has been synthesized based on our results described in section 5.1. We note that this table does not include those behaviors that stayed consistent for a certain group of learners between the two strategies. For example, for *Silent Wanderers*, the fact that we do not see negative affect in the table indicates that there were not any significant oscillations for their negative valence between the two strategies, i.e., their negative emotions were more consistent irrespective of which problem strategy they used.

When doing global problem solving consisting predominantly of additions, the two gainer groups *Expressive Explorers* and *Calm Tinkerers* have high speech, while *Silent Wanderers* speak relatively less. In this phase, the two gainer groups gaze at their screen is high, while the gaze towards their partner or the robot is lowest. On the other hand, for the non-gainer group *Silent Wanderers*, while the gaze towards the screen is

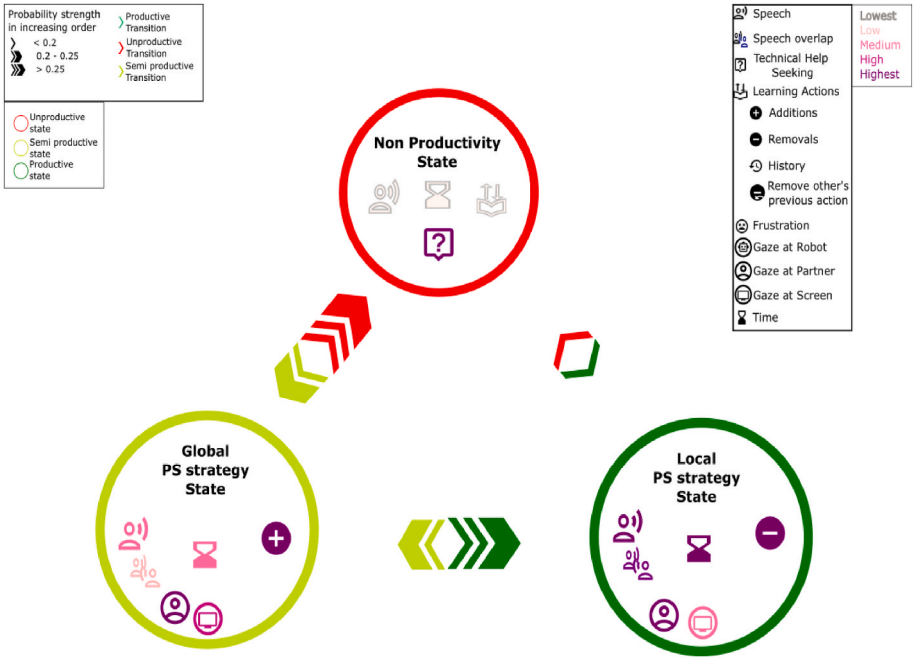


Fig. 10. Temporal profile for Silent Wanderers.

Table 4
Interplay between stages of problem solving strategies and behaviors of speech, gaze, and affect.

When employing a global problem solving strategy			
Behavior	Expressive Explorers	Calm Tinkerers	Silent Wanderers
Speech	High	High	Medium
Gaze towards partner and/ or robot	Lowest	Lowest	Highest
Gaze towards the screen	High	High	High
Affect	Medium Negative	Highest both	Highest Positive
When employing a local problem solving strategy			
Behavior	Expressive Explorers	Calm Tinkerers	Silent Wanderers
Speech	Highest	Highest	Highest
Gaze towards partner and/ or robot	Highest	High	Medium
Gaze towards the screen	Lowest	Medium	Medium
Affect	Highest Negative	Medium both	High Positive

high, their gaze towards their partner is highest in this phase. Lastly, in terms of affect, *Expressive Explorers* express medium level of negative emotions, *Calm Tinkerers* display both highest levels of positive as well as negative emotions in this phase, while the non-gainer group *Silent Wanderers* are associated with their highest levels of positive emotions in this phase.

Next, we observe that when using the local problem solving strategy, i.e., more removals, an action indicative of reflection, each group's *speech activity* is at their highest. In terms of gaze behavior, the two gainer groups *Expressive Explorers* and *Calm Tinkerers* gaze at their partners as well as the robot is high in this phase, while *Silent Wanderers* gaze towards their partner is lesser. Furthermore, *Expressive Explorers* gaze towards the screen is the lowest in this phase, while the other two groups gaze at the screen is medium. Lastly, *Expressive Explorers* show most negative emotions during this strategy, *Calm Tinkerers* are associated with medium emotions, while *Silent Wanderers* lean towards high positive emotions while removing.

It is interesting to note that irrespective of the phase of problem solving, both gainer groups maintain a high level of verbal interaction as opposed to the non-gainer group *Silent Wanderers* who speak less during global problem solving and speak the most while in the local problem solving phase. This suggests that verbal interactions are important to be maintained during both the global and local problem solving phases, i.e. both when making additions, as well as when doing removals. The need for communication itself is not surprising as the collaborative problem solving task requires learners to share information for building a common ground and improving their understanding to construct a solution, monitor and reflect on the solution (Barron, 2003; Chang et al., 2017; Hausmann et al., 2004; Roschelle & Teasley, 1995). Our analysis reiterates the need for communication throughout collaborative problem solving, regardless of the PS strategy being applied. Nevertheless some phases may demand a higher level of interaction between partners. For instance, literature suggests an increase in interaction between participants during phases of socially shared regulation of learning which involves reflection, monitoring the solution that has been built and evaluating whether to revise it (Isohätälä et al., 2017; Rogat & Linnenbrink-Garcia, 2011; Sinha et al., 2015). We also find similar behaviors in that we see an increase in speech activity of all learners in their most reflective phase of problem solving, which in our case is the local problem solving that involves continuously evaluating whether an added edge satisfies the requirement of minimising cost and removing it if not. This requires partners to share the information on their respective screens and discuss it with respect to the overall solution, thus leading to increase in speech.

In terms of affect, all groups oscillate between different affective states and/or different *levels of affect*. *Expressive Explorers* oscillate between medium and very high negative valence levels during global and local phases respectively, i.e., *showing a higher frustration during the local strategy*. On the other hand, the second type of gainers, *Calm Tinkerers* oscillate between higher to medium level of arousal, with a mix of both positive and negative valence, when moving respectively between global and local problem solving, i.e, *displaying higher levels of both excitement and frustration during the global strategy*. Lastly, for *Silent Wanderers*, the oscillation is more in terms of arousal, that shifts between their relative levels of highest to high positive valence between global and local problem solving, respectively, i.e, *being more excited during global problem solving*. The changing dynamics of affective states over the entire problem solving is supported by the work of D'Mello and Graesser (2012); however, what is interesting is that *both gainer groups experience negative emotions during both global and local problem solving phases*. A meta-analysis of discrete affective states during learning with technology indicates that negative states such as anger, contempt, sadness, anxiety, fear, etc. are relatively infrequently experienced when students engage with technology-enhanced learning contexts (D'Mello, 2013). However, these learning contexts are guided discovery learning contexts that usually employ success-driven scaffolding to nudge the learners towards the correct solution. Sinha (2021b), in a recent work suggested that in a problem-solving followed by instruction (PS-I) context, where the problem-solving phase is "naturally designed to be ill-structured and afford the generation of multiple suboptimal solutions (Kapur & Bielaczyc, 2012)", some levels of negative emotions can in fact be beneficial as they can "keeps one alerted of challenges requiring more focused attention, and assists in comprehending conflicting information (Ivtzan et al., 2015; Kashdan & Biswas-Diener, 2014)". Since our open-ended activity is also designed as a PS-I activity, the surfacing of absolute medium levels of negative emotions among gainers (the mean values can be seen in the Tables in the appendix; note that the labels highest, high, medium, low, lowest are relative within a group) can be considered as supporting what was reported in Sinha (2021b). In this work, we additionally point out *when* negative emotions increase during problem-solving, relative to other phases.

Another point of interest is that while the interplay between problem solving strategy and affect was highlighted in our previous work (Nasir et al., 2021c), this work highlights that a particular affect is not strictly associated with a *type* of problem solving strategy but it also depends on the phase of the activity and a particular problem solving strategy applied at the later stages of the activity can lead to more negative emotions than would be otherwise observed. In D'Mello and Graesser (2012), the authors highlight that moving from a state of equilibrium or flow to a state of disequilibrium results in negative emotions such as confusion and frustration. Our findings of gainers emotions also suggests a similar behavior; for instance when *Expressive Explorers* change strategies from a global to a local one, it is accompanied by an increase in negative emotions and when *Calm Tinkerers* shift from a local to a global strategy they show an increase in negative emotions. This *change* in negative emotions in not very prominent among *Silent Wanderers* which could be because they did not pay as much attention to the task at hand or notice the gaps in their prior knowledge and the need for reflection (Sinha, 2021b).

Oscillation of gaze between the partner and the screen, and the robot and the screen, is particularly interesting as we observe that for both gainer groups, they look the least at their partner or at the robot when employing the global PS strategy but highest during the local PS strategy. On the contrary, the non-gainer group looks more to their partner and the robot when exhibiting global PS strategy compared to the local PS strategy. Literature suggests that gaze is a means of action monitoring, predicting intention, action co-ordination and planning in order to establish a common ground that can lead to better collaboration (Huang et al., 2015; Sebanz et al., 2006). Together our findings and literature suggest that in an environment that has both social (a partner) and task elements (screens), looking at your partner during the local PS

strategy, which involves mostly removing what the team has already built and requires agreeing on which edges to remove, can support joint action. Since in this work we do not distinguish between moments when both partners are looking at each other and when one partner is looking at the other (both are considered when computing the feature "gaze at partner"), eye gaze could either be a way to confirm agreement on a bilaterally decided course of action or a way to negotiate to reach a consensus when a unilateral decision was taken. On the other hand, during the global PS strategy which involves series of additions, it is more productive to look at the screen rather than at the partner as the plan is already agreed on (global reflection/planning).

5.3. Connections to computer-supported collaborative learning literature

Within CSCL literature the temporal analysis of computer-supported collaborative learning (Lämsä et al., 2021) has predominantly focussed on the content of learners verbal communication/interaction/discussion and how it evolves during the learning activity, with the non-verbal activities such as actions within the technology-based learning environment, serving to complement the analysis of verbal communication. In our work, we employ multimodal features to understand how pairs of students learn by working on an open-ended scripted collaborative problem-solving activity. For this, we consider the pair as a single unit and examine how their collective behaviors (speech activity, problem-solving actions, eye gaze and affect) change across the activity as they learn by problem-solving. Our analysis does not include any measure of the quality of the verbal discussion, but studies the temporal evolution of this units' learning behaviors using only fully quantitative data and methods. Similar methods have been used in (Martinez-Maldonado et al., 2013) where the authors were able to distinguish between high and low collaborating groups based on their action and speech sequences and our work adds to this literature by additionally considering affect and eye gaze, and modeling the temporal learning process of different types of learners.

Further, using the quality of speech, with and without problem-solving actions, has allowed researchers to understand how learners temporally regulate their open-ended problem-solving (Chang et al., 2017; Emara et al., 2021; Kapur, 2011; Malmberg et al., 2015; Sobocinski et al., 2017) in face-to-face collaborative conditions. For instance, researchers identified that increased socially shared regulation across time corresponded with increased use of more systematic action sequences (Emara et al., 2021) and higher performance (Malmberg et al., 2015). Similarly, Sobocinski et al. (2017) found that in low challenge sessions, learners transitioned between the forethought and performance phases of self-regulated learning only once, while in high challenge sessions they transitioned between forethought and performance phases more frequently. Chang et al. (2017) identified that successful groups discourse transitioned more frequently from monitoring to formulating and exploring, along with doing exploratory actions, as opposed to less successful groups whose discourse suggested a more trial-and-error strategy. While we did not explicitly identify socially shared regulation, our findings did agree with the above findings in that increased speech activity was overall associated with increased reflective problem-solving actions, both global and local. In addition, our work offers a complementary view of how collaborative open-ended problem-solving proceeds, in terms of problem-solving strategies (local vs global) rather than problem-solving phases (exploring, formulating, planning and monitoring). The global problem solving strategy can be considered as one in which planning, exploring, formulating and monitoring happens on the scale of the entire problem. The local problem solving strategy is one in which the planning, exploring, formulating and monitoring happens on the scale of the next step towards the solution. Our work thus adds to CSCL literature by suggesting that learners seamlessly intertwine these two strategies in their productive collaborative problem-solving, and that neither is at the outset "*better*" than the other.

5.4. Implications for design of adaptive learning interventions

In this subsection, we highlight some implications of the findings discussed above for the design of adaptive learning interventions, both at a broader level for the CSCL community, and at the specific level of the intervention in our study. To summarize our observations from the temporal profiles, we find that:

1. All learner groups have the highest probability to start with and stay in a state of non-productivity. However, once out of it, all learners have the lowest probability to return to this state.
2. The non-gainers transition between states of non-productivity and productivity in a smoother manner with an intermediate semi-productive state in terms of time. In contrast, gainers' transitions are sharper, in that they transition from the non-productive state to one of the two productive states.

3. *Expressive Explorers* and *Calm Tinkerers* do not exclusively adopt a global or a local PS approach respectively throughout the activity, as suggested by the aggregate behavioral profiles in Nasir et al. (2021c). This analysis reveals that both these gainer types adopt both these approaches and switch between them throughout the interaction. One key difference is the stages of the interaction in which the two groups employ the strategies, with the *Expressive Explorers* adopting the global strategy earlier and then the local strategy, while the *Calm Tinkerers* adopting the reverse approach.

4. Further, for the two gainer groups, each of the two problem solving strategies is associated with speech, gaze and affect in a unique way, that is in some ways comparable (speech and gaze) and in other ways opposing (affect). Diving deeper, the relationship of affect with a particular problem solving strategy does not seem to be as straightforward as suggested by aggregate behavioral analysis in .Nasir et al. (2021c) Both types of gainers seem to have increased emotional behavior relative to themselves towards the later part of the interaction irrespective of which problem solving strategy they are using.

Following up from the above observations, (1) suggests that adaptive interventions should start early in the interaction, irrespective of the group. For example, all groups speak the least in the non-productive state and have yet not established either of the problem solving strategies. An effective intervention could then be to try to induce communication between the dyad earlier in the interaction, that eventually could help with mitigating confusion, building a common ground, resolving conflict and pushing the team towards a more reflective set of behaviors, i.e., to follow either a global or local problem solving strategy.

Further, going back more often (i.e., with a higher probability) into a non-productive state of low speech (as *Silent Wanderers* as well as *Calm Tinkerers* did) might suggest that the students have not yet established a shared understanding of the problem. Without an appropriate intervention, the relevant team may take longer to have productive interactions or transition to a productive state. Such an unstable behavior of moving back and forth between the non-productive and productive states need to be mitigated by an intervention targeted at inducing behaviors that would increase the chances of building a shared understanding. Further, observation of *Silent Wanderers* suggests that it is the lack of reflective actions such as looking back at their previous solutions and observing their own or their partners action, that might be the cause of a delayed shared understanding of the problem. Hence, such actions can be additionally suggested by an intelligent agent if the team is observed to be going back often to a state of lower speech that suggests being in a non-productive state.

Lastly, as highlighted by (3) and (4), identification of a team as following a local or global PS strategy at the early stages of the interaction should be taken with caution. Instead continuous identification of the teams current PS strategy is necessary as the teams shift between multiple PS strategies and each problem solving strategy elicits different

speech, gaze and affective behavior in learners. Therefore, it is important to inform the mechanism behind interventions of this sophisticated interplay and suggest interventions accordingly. For example, *Expressive Explorers* increase in their intensity of negative emotions as they move from global to local PS strategy and vice versa for the *Calm Tinkerers*; however, when looking at the time axis, in both cases this increase is towards the later phase of the interaction. Hence, the adaptive intervention system does not always need to mitigate frustration, especially towards the end of the interaction as this level of frustration may be conducive to more productive behaviors. This can be an interesting avenue for further investigation by the community. As another example, both gainer groups looking more at the partner when moving from global to local PS strategy seems to suggest better collaboration quality; therefore, the adaptive intervention system can try to induce relevant gaze behaviors when the associated PS strategy is detected among learners potentially by sharing gaze among the peers as has been shown to be effective Schneider et al. (2018).

6. Conclusion

Concluding on our discussion, in this paper we contribute by applying an HMM based methodology to model and understand the *collaborative learning process* of gainer and non-gainer teams. However, there are some limitations with the current study. Firstly, in order to generalize the outcomes and inferences to collaborative settings in open-ended environments, there is a need of carrying out even more extensive studies, i.e., with more teams. Then, the current data is skewed when it comes to non-gainer teams, that is we have lesser non-gainer teams in our data than gainer teams and that can add to making our results less straightforward to generalize. Lastly, since the study is done at international schools in Switzerland, the students are from a selective pool coming from a certain economic and social background; hence, this requires us to be careful about the group we generalize it to.

In our future work, our goal is to use these findings to build an adaptive intervention mechanism for a robot that can observe the multimodal behaviors of the students in soft real-time and provide effective interventions. With such a robot, we plan to collect more data to account for the aforementioned limitations, by both testing the effectiveness of our adaptive system, refining it as well as observing if the new data generalizes to similar (and even additional) profiles.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955.

Appendix A

Table 5
Features' Mean values in each of the Expressive Explorers' states

Feature	InitialState	MoreProbableState	LessProbableState
T_add	1.102384×10^{-9}	3.573773×10^{-1}	0.040838
T_ratio_add_rem	4.183066×10^{-9}	9.9999931×10^{-1}	0.018960
T_action	3.870495×10^{-2}	9.699460×10^{-2}	0.044758
normalized_time	2.029434×10^{-1}	5.387081×10^{-1}	0.626040
Speech_Overlap	2.723118×10^{-1}	4.758125×10^{-1}	0.567795
Overlap_to_Speech_Ratio	5.461976×10^{-1}	6.965724×10^{-1}	0.804408
Speech_Activity	4.099696×10^{-1}	5.986744×10^{-1}	0.665616
Silence	6.541648×10^{-1}	4.866740×10^{-1}	0.410024
T_remove	9.325293×10^{-10}	3.182062×10^{-7}	0.126406
Gaze_at_Robot	4.343753×10^{-2}	9.518764×10^{-3}	0.045563
redundant_exist	3.763263×10^{-3}	6.830674×10^{-3}	0.002477
T1_T1_rem	1.027991×10^{-17}	3.735072×10^{-20}	0.116683
Gaze_at_Partner	7.156486×10^{-2}	6.737566×10^{-2}	0.117361
T_help	7.807358×10^{-2}	6.557370×10^{-3}	0.014712
T1_T2_rem	1.478677×10^{-15}	4.773094×10^{-7}	0.043755
T_hist	5.047627×10^{-3}	5.044131×10^{-3}	0.001290
Gaze_at_Screen_Right	5.915012×10^{-1}	5.912295×10^{-1}	0.585986
Gaze_at_Screen_Left	3.447677×10^{-1}	3.441521×10^{-1}	0.306201
Long_Pauses	4.414569×10^{-3}	1.723356×10^{-2}	0.002917
Arousal	2.705875×10^{-1}	3.101827×10^{-1}	0.375027
Short_Pauses	1.685203×10^{-1}	1.542912×10^{-1}	0.116228
Negative_Valence	2.056995×10^{-1}	2.568086×10^{-1}	0.308619
Positive_Valence	3.375673×10^{-1}	3.469566×10^{-1}	0.412408
Gaze_Other	8.812433×10^{-2}	5.841153×10^{-2}	0.062550
T1_T2_add	0.000000	0.000000	0.000000
Difference_in_Valence	5.507043×10^{-1}	5.013340×10^{-1}	0.513887
T1_T1_add	0.000000	0.000000	0.000000

Table 6
Features' Mean values in each of the Calm Tinkerers' states

Feature	InitialState	MoreProbableState	LessProbableState
T_ratio_add_rem	2.673037×10^{-10}	1.000 000	4.172444×10^{-3}
T_add	6.682594×10^{-10}	3.166667×10^{-1}	1.043111×10^{-2}
Speech_Overlap	3.282486×10^{-1}	5.413534×10^{-1}	5.921923×10^{-1}
Speech_Activity	4.855440×10^{-1}	6.827465×10^{-1}	7.096177×10^{-1}
Silence	5.495654×10^{-1}	3.864979×10^{-1}	3.520730×10^{-1}
T_action	1.260936×10^{-2}	5.866667×10^{-2}	2.662700×10^{-2}
Overlap_to_Speech_Ratio	6.216050×10^{-1}	7.520540×10^{-1}	7.951161×10^{-1}
normalized_time	3.205108×10^{-1}	5.447131×10^{-1}	5.415407×10^{-1}
T_remove	1.266331×10^{-2}	3.999698×10^{-15}	1.701398×10^{-2}
T1_T1_rem	1.168974×10^{-12}	8.479291×10^{-18}	6.258666×10^{-2}
T1_T2_rem	1.347584×10^{-7}	2.299210×10^{-21}	2.086218×10^{-2}
redundant_exist	1.747415×10^{-3}	1.458333×10^{-2}	5.336480×10^{-3}
Positive_Valence	3.665011×10^{-1}	4.497916×10^{-1}	4.121070×10^{-1}
Arousal	3.269324×10^{-1}	3.862492×10^{-1}	3.702702×10^{-1}
Gaze_at_Robot	1.162740×10^{-2}	5.383023×10^{-3}	1.723160×10^{-2}
Negative_Valence	2.549870×10^{-1}	2.911126×10^{-1}	2.905847×10^{-1}
T_help	1.132622×10^{-2}	7.855360×10^{-22}	1.394327×10^{-2}
Gaze_at_Screen_Right	5.173883×10^{-1}	5.228333×10^{-1}	4.819030×10^{-1}
Short_Pauses	6.129101×10^{-2}	6.038230×10^{-2}	5.266563×10^{-2}
Difference_in_Valence	5.511903×10^{-1}	6.026717×10^{-1}	5.586536×10^{-1}
Gaze_at_Partner	1.790690×10^{-1}	1.355978×10^{-1}	1.555657×10^{-1}
Gaze_at_Screen_Left	4.294050×10^{-1}	4.452297×10^{-1}	4.510707×10^{-1}
Long_Pauses	1.474565×10^{-2}	9.933266×10^{-3}	1.916058×10^{-3}
T1_T2_add	3.027555×10^{-32}	1.666667×10^{-2}	9.423054×10^{-19}
Gaze_Other	5.388054×10^{-2}	7.411003×10^{-2}	6.656883×10^{-2}
T_hist	9.891350×10^{-3}	8.333333×10^{-3}	2.176802×10^{-2}
T1_T1_add	0.000 000	0.000 000	0.000 000

Table 7
Features' Mean values in each of the Silent Wanderers' states

Feature	InitialState	MoreProbableState	LessProbableState
T_ratio_add_rem	1.312014×10^{-2}	9.999978×10^{-1}	6.729559×10^{-3}
T_add	3.644306×10^{-2}	3.281263×10^{-1}	1.682390×10^{-2}
Speech_Overlap	6.135501×10^{-2}	1.679682×10^{-1}	4.213296×10^{-1}
Speech_Activity	2.082582×10^{-1}	3.460734×10^{-1}	5.755465×10^{-1}
Overlap_to_Speech_Ratio	1.891342×10^{-1}	3.170220×10^{-1}	6.084412×10^{-1}
Silence	7.586753×10^{-1}	6.372338×10^{-1}	4.604999×10^{-1}
T_action	5.638728×10^{-2}	1.191416×10^{-1}	5.272196×10^{-2}
normalized_time	3.157809×10^{-1}	4.738481×10^{-1}	7.326249×10^{-1}
T_remove	1.092886×10^{-1}	1.552110×10^{-6}	1.347397×10^{-1}
redundant_exist	3.997135×10^{-2}	5.468935×10^{-2}	2.257087×10^{-2}
Gaze_at_Screen_Right	5.556238×10^{-1}	6.204418×10^{-1}	6.182313×10^{-1}
Gaze_at_Screen_Left	2.746859×10^{-1}	2.511985×10^{-1}	2.227564×10^{-1}
Positive_Valence	2.501105×10^{-1}	2.826254×10^{-1}	2.784481×10^{-1}
T_help	3.497000×10^{-2}	6.249981×10^{-3}	1.707482×10^{-10}
Gaze_at_Partner	1.141224×10^{-1}	1.443724×10^{-1}	1.290232×10^{-1}
Difference_in_Valence	3.721043×10^{-1}	3.680229×10^{-1}	3.831570×10^{-1}
Arousal	2.464960×10^{-1}	3.058044×10^{-1}	2.857545×10^{-1}
T1_T1_rem	2.914167×10^{-2}	1.625856×10^{-13}	1.646490×10^{-12}
T1_T2_rem	6.827181×10^{-5}	4.079701×10^{-11}	3.360833×10^{-2}
Gaze_Other	4.831607×10^{-2}	1.046721×10^{-1}	5.910927×10^{-2}
Gaze_at_Robot	8.571055×10^{-2}	4.328244×10^{-2}	4.276117×10^{-2}
T1_T1_add	0.000 000	0.000 000	0.000 000
Negative_Valence	2.408355×10^{-1}	3.125240×10^{-1}	2.792128×10^{-1}
Short_Pauses	2.230099×10^{-1}	1.495351×10^{-1}	1.287077×10^{-1}
T1_T2_add	5.158890×10^{-15}	6.249980×10^{-2}	6.094868×10^{-10}
Long_Pauses	2.606827×10^{-2}	6.442913×10^{-3}	9.601131×10^{-3}
T_hist	7.199166×10^{-3}	2.083376×10^{-2}	3.378669×10^{-2}

Table 8
p-values from Kruskal-Wallis test on the Expressive Explorers' states

Feature	LessProbableState-MoreProbableState	LessProbableState-InitialState	MoreProbableState-InitialState	LessProbableState-MoreProbableState-InitialState

(*continued on next page*)

Table 8 (*continued*)

Feature	LessProbableState-MoreProbableState	LessProbableState-InitialState	MoreProbableState-InitialState	LessProbableState-MoreProbableState-InitialState
T_add	$6.459889 \times 10^{-272}$	7.786821×10^{-7}	$7.321670 \times 10^{-241}$	0.000 000
T_ratio_add_rem	0.000 000	8.011061×10^{-7}	$6.043280 \times 10^{-301}$	0.000 000
T_action	3.229241×10^{-62}	1.116775×10^{-12}	$1.255941 \times 10^{-12}8$	$2.724522 \times 10^{-13}4$
normalized_time	6.573843×10^{-13}	$1.588430 \times 10^{-11}8$	2.121637×10^{-76}	$2.268858 \times 10^{-12}7$
Speech_Overlap	1.823268×10^{-20}	4.726497×10^{-96}	4.636181×10^{-33}	1.145578×10^{-91}
Overlap_to_Speech_Ratio	2.359913×10^{-16}	4.714581×10^{-82}	2.061077×10^{-29}	8.026889×10^{-78}
Speech_Activity	1.878127×10^{-18}	1.021272×10^{-76}	9.454754×10^{-27}	1.719850×10^{-74}
Silence	4.052784×10^{-11}	1.423945×10^{-69}	4.624967×10^{-33}	8.671185×10^{-69}
T_remove	1.923088×10^{-30}	8.584127×10^{-11}	3.403098×10^{-8}	1.086990×10^{-34}
Gaze_at_Robot	2.408710×10^{-20}	2.522376×10^{-1}	1.119077×10^{-13}	4.107882×10^{-21}
redundant_exist	6.323133×10^{-13}	5.619846×10^{-1}	5.410339×10^{-12}	5.495916×10^{-18}
T1_T1_rem	1.071271×10^{-9}	1.175112×10^{-6}	NaN	7.707447×10^{-14}
Gaze_at_Partner	8.078957×10^{-11}	4.279389×10^{-8}	8.546495×10^{-1}	1.426036×10^{-11}
T_help	5.077449×10^{-2}	1.210030×10^{-5}	5.427370×10^{-10}	1.167470×10^{-10}
T1_T2_rem	3.871818×10^{-7}	6.822129×10^{-4}	7.513610×10^{-2}	2.112531×10^{-8}
T_hist	9.259736×10^{-6}	2.101519×10^{-6}	3.011715×10^{-1}	1.159495×10^{-7}
Gaze_at_Screen_Right	3.147718×10^{-7}	6.138099×10^{-2}	1.809002×10^{-3}	8.358584×10^{-7}
Gaze_at_Screen_Left	5.380571×10^{-6}	3.960269×10^{-4}	6.152767×10^{-1}	8.665531×10^{-6}
Long_Pauses	3.314372×10^{-4}	1.312236×10^{-3}	9.461985×10^{-1}	4.948046×10^{-4}
Arousal	4.132118×10^{-3}	4.371238×10^{-4}	3.136678×10^{-1}	7.750580×10^{-4}
Short_Pauses	1.013922×10^{-2}	2.814769×10^{-4}	1.729218×10^{-1}	8.445037×10^{-4}
Negative_Valence	3.444524×10^{-3}	5.710819×10^{-3}	8.727249×10^{-1}	3.942160×10^{-3}
Positive_Valence	1.202900×10^{-1}	8.595192×10^{-3}	1.684401×10^{-1}	2.711117×10^{-2}
Gaze_Other	6.718909×10^{-1}	6.410131×10^{-2}	1.878263×10^{-2}	5.268550×10^{-2}
T1_T2_add	1.782952×10^{-1}	NaN	2.607401×10^{-1}	2.148112×10^{-1}
Difference_in_Valence	1.769718×10^{-1}	7.301056×10^{-1}	3.636468×10^{-1}	3.731152×10^{-1}
T1_T1_add	7.361626×10^{-1}	2.372005×10^{-1}	1.682336×10^{-1}	4.040213×10^{-1}

Table 9
p-values from Kruskal-Wallis test on the Calm Tinkerers' states

Feature	InitialState-MoreProbableState	InitialState-LessProbableState	MoreProbableState-LessProbableState	InitialState-MoreProbableState-LessProbableState
T_ratio_add_rem	$1.046633 \times 10^{-214}$	3.203160×10^{-15}	$1.448010 \times 10^{-209}$	$2.208200 \times 10^{-300}$
T_add	$8.825740 \times 10^{-194}$	2.671724×10^{-15}	$1.084962 \times 10^{-13}6$	$2.779006 \times 10^{-241}$
Speech_Overlap	2.486453×10^{-26}	$1.140660 \times 10^{-10}5$	4.815798×10^{-22}	$3.079549 \times 10^{-10}3$
Speech_Activity	2.699760×10^{-26}	4.539962×10^{-96}	4.068440×10^{-19}	1.789244×10^{-94}
Silence	6.369222×10^{-31}	1.859685×10^{-91}	9.868460×10^{-16}	7.74×10^{-92}
T_action	2.530858×10^{-96}	9.718553×10^{-15}	3.560970×10^{-28}	2.236893×10^{-84}
Overlap_to_speech_ratio	1.266769×10^{-19}	1.470592×10^{-84}	9.053385×10^{-20}	1.396038×10^{-82}
Normalized_time	9.962378×10^{-23}	1.277583×10^{-57}	2.040765×10^{-10}	3.391638×10^{-59}
T_remove	4.195097×10^{-14}	3.443724×10^{-19}	4.953212×10^{-44}	3.040368×10^{-51}
T1_T1_rem	3.566311×10^{-1}	6.479934×10^{-18}	4.254322×10^{-16}	2.402922×10^{-30}
T1_T2_rem	1.921375×10^{-1}	1.517800×10^{-9}	2.073554×10^{-9}	6.086552×10^{-16}
Redundant_exist	7.350425×10^{-13}	2.140998×10^{-2}	2.082055×10^{-7}	1.988810×10^{-13}
Positive_Valence	1.140000×10^{-4}	8.498930×10^{-11}	2.452327×10^{-2}	5.318031×10^{-10}
Arousal	5.355177×10^{-2}	1.451686×10^{-9}	1.509734×10^{-4}	5.760692×10^{-9}
Gaze_at_robot	9.302748×10^{-8}	4.740781×10^{-3}	4.274607×10^{-3}	5.047403×10^{-7}
Negative_Valence	7.412760×10^{-1}	3.148931×10^{-6}	3.576009×10^{-5}	1.455941×10^{-6}
T_help	6.274225×10^{-5}	2.649523×10^{-4}	3.834952×10^{-1}	5.631461×10^{-6}
Gaze_at_screen_right	3.407637×10^{-4}	9.740129×10^{-2}	3.593597×10^{-6}	5.796425×10^{-6}
Short_pauses	4.721115×10^{-3}	9.175213×10^{-7}	9.540181×10^{-2}	6.114326×10^{-6}
Difference_in_Valence	4.211367×10^{-6}	6.285657×10^{-3}	5.459578×10^{-2}	2.763204×10^{-5}
Gaze_at_partner	1.282025×10^{-1}	6.511073×10^{-5}	2.696452×10^{-2}	3.242851×10^{-4}
Gaze_at_screen_left	1.130881×10^{-3}	4.568822×10^{-1}	1.584547×10^{-2}	4.372545×10^{-3}
Long_pauses	6.461623×10^{-1}	1.588477×10^{-2}	3.468741×10^{-3}	8.647671×10^{-3}
T1_T2_add	4.872548×10^{-3}	2.015286×10^{-2}	5.295830×10^{-1}	2.237594×10^{-2}
Gaze_other	8.545218×10^{-1}	8.659117×10^{-2}	1.429300×10^{-1}	1.664223×10^{-1}
T_hist	7.147372×10^{-1}	3.024626×10^{-1}	1.804699×10^{-1}	3.499533×10^{-1}
T1_T1_add	8.703039×10^{-1}	5.027353×10^{-1}	4.142353×10^{-1}	7.077864×10^{-1}

Table 10
p-values from Kruskal-Wallis test on the Silent Wanderers' states

Feature	InitialState-LessProbableState	InitialState-MoreProbableState	LessProbableState-MoreProbableState	InitialState-LessProbableState-MoreProbableState
T_ratio_add_rem	4.589301×10^{-1}	$5.220515 \times 10^{-12}2$	$6.663973 \times 10^{-13}0$	$5.535321 \times 10^{-18}3$
T_add	4.577846×10^{-1}	$6.010859 \times 10^{-10}0$	$2.610590 \times 10^{-11}4$	$2.388349 \times 10^{-16}0$
Speech_overlap	5.892624×10^{-97}	7.100573×10^{-18}	9.712381×10^{-31}	2.760344×10^{-98}
Speech_Activity	8.477783×10^{-90}	2.493802×10^{-20}	7.562644×10^{-23}	1.735427×10^{-89}
Overlap_to_speech_ratio	7.595498×10^{-72}	3.505213×10^{-11}	6.190938×10^{-30}	1.645284×10^{-75}
Silence	3.187753×10^{-70}	1.142510×10^{-13}	2.080269×10^{-19}	4.649497×10^{-69}
T_action	1.056353×10^{-1}	3.188625×10^{-32}	1.079551×10^{-44}	3.200090×10^{-49}
Normalized_time	1.518080×10^{-46}	3.510905×10^{-9}	5.506350×10^{-15}	2.159810×10^{-46}
T_remove	6.971572×10^{-1}	1.116435×10^{-11}	6.037973×10^{-11}	1.127712×10^{-10}
Redundant_exist	3.512804×10^{-1}	1.635185×10^{-6}	3.081867×10^{-9}	1.046462×10^{-9}
Gaze_at_screen_right	4.238118×10^{-1}	2.854616×10^{-7}	1.300446×10^{-5}	4.386002×10^{-7}
Gaze_at_screen_left	1.959077×10^{-4}	2.097955×10^{-6}	1.816661×10^{-1}	3.48×10^{-6}
Positive_Valence	4.477664×10^{-5}	8.188481×10^{-1}	3.719569×10^{-5}	8.771086×10^{-6}
T_help	2.145970×10^{-4}	4.619297×10^{-3}	5.533612×10^{-1}	1.305032×10^{-4}
Gaze_at_partner	6.836966×10^{-5}	3.239169×10^{-1}	8.317920×10^{-3}	2.405678×10^{-4}
Difference_in_Valence	5.798869×10^{-4}	7.765849×10^{-1}	4.280245×10^{-3}	9.162789×10^{-4}
Arousal	3.133786×10^{-2}	2.050905×10^{-1}	4.771730×10^{-4}	2.053134×10^{-3}
T1_T1_rem	2.168625×10^{-1}	4.151916×10^{-3}	3.627620×10^{-2}	1.568344×10^{-2}
T1_T2_rem	6.483012×10^{-1}	1.048949×10^{-2}	4.416490×10^{-3}	2.137952×10^{-2}
Gaze_other	1.069325×10^{-2}	4.460858×10^{-2}	8.001324×10^{-1}	2.297604×10^{-2}
Gaze_at_robot	1.107108×10^{-1}	8.746928×10^{-3}	2.324421×10^{-1}	2.894125×10^{-2}
T1_T1_add	NaN	1.155100×10^{-1}	9.639491×10^{-2}	7.286920×10^{-2}
Negative_Valence	9.084048×10^{-1}	6.888763×10^{-2}	3.284085×10^{-2}	8.004833×10^{-2}
Short_pauses	3.175136×10^{-1}	9.025873×10^{-2}	2.226728×10^{-1}	1.799852×10^{-1}
T1_T2_add	5.001169×10^{-1}	2.773726×10^{-1}	8.430488×10^{-2}	1.889137×10^{-1}
Long_pauses	6.990590×10^{-1}	1.992827×10^{-1}	2.787486×10^{-1}	3.964593×10^{-1}
T_hist	3.746288×10^{-1}	9.370430×10^{-1}	3.645077×10^{-1}	5.606479×10^{-1}

References

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning, 9*, 161–185. https://doi.org/10.1007/s11409-013-9107-6, 10.1007/s11409-013-9107-6.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307–359.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*, Article eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Blikstein, P. (2013). *Multimodal learning analytics*. https://doi.org/10.1145/2460296.2460316

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*, 220–238. https://doi.org/10.18608/jla.2016.32.11

Brooks, J. M., & Brooks, M. (1993). In *Search of understanding: The case for constructivist classrooms*.

Castanheira, M. L., Crawford, T., Dixon, C. N., & Green, J. L. (2000). Interactional ethnography: An approach to studying the social construction of literate practices. *Linguistics and Education, 11*, 353–400. https://doi.org/10.1016/S0898-5898(00)00032-2

Chalmers, C. (2018). Robotics and computational thinking in primary school. *IJCCI, 17*, 93–100.

Chang, C.-j., Chang, M.-h., Chiu, B.-c., Liu, C.-c., Chao, P.-y., Lai, C.-h., Wu, S.-w., Chang, C.-k., & Chen, W. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education, 114*, 222–235. https://doi.org/10.1016/j.compedu.2017.07.008, 10.1016/j.compedu.2017.07.008.

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100002. https://doi.org/10.1016/j.caeai.2020.100002. URL: https://www.sciencedirect.com/science/article/pii/S2666920X20300023.

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. *Educational Technology & Society, 25*, 28–47. URL: https://www.jstor.org/stable/48647028.

Chow, J. Y., Davids, K., Button, C., & Renshaw, I. (2015). *Nonlinear pedagogy in skill acquisition: An introduction*. Routledge.

Ciasullo, M. V., Carli, M., Lim, W. M., & Palumbo, R. (2022). An open innovation approach to co-produce scientific knowledge: An examination of citizen science in the healthcare ecosystem. *European Journal of Innovation Management, 25*, 365–392. https://doi.org/10.1108/EJIM-02-2021-0109. URL: https://www.ingentaconnect.com/content/mcb/220/2021/00000025/00000006/art00016.

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253–278. https://doi.org/10.1007/BF01099821, 10.1007/BF01099821.

Csanadi, A., Eagan, B. R., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ena) to analyze verbal data in cscl research. *International Journal of Computer-Supported Collaborative Learning, 13*, 419–438.

Desmarais, M. C., & Baker, R. S. J.d. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction, 22*, 9–38. https://doi.org/10.1007/s11257-011-9106-8, 10.1007/s11257-011-9106-8.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105*, 1082. https://doi.org/10.1037/a0032674

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*, 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001, 10.1016/j.learninstruc.2011.10.001.

Do-lenh, S. (2012). *Supporting reflection and classroom orchestration with tangible tabletops* (Vol. 5313, p. 241). https://doi.org/10.5075/epfl-thesis-5313

Emara, M., Hutchins, N., Grover, S., Snyder, C., & Biswas, G. (2021). Examining student regulation of collaborative, computational, problem-solving processes in open-ended learning environments. *Journal of Learning Analytics, 8*, 49–74. https://doi.org/10.18608/jla.2021.7230

Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, 51.

Engelmann, K., & Bannert, M. (2021). *Analyzing temporal data for understanding the learning process induced by metacognitive prompts* (Vol. 72), Article 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management, 48*, 108–119. https://doi.org/10.1016/j.ijinfomgt.2019.02.003. URL: https://www.sciencedirect.com/science/article/pii/S0268401218312751.

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., Das, M., & Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the 30th conference on artificial intelligence* (pp. 3951–3957). AAAI 2016.

Guo, T., Lin, T., & Antulov-Fantulin, N. (2019). *Exploring interpretable lstm neural networks over multi-variable data*. arXiv:1905.12034.

Hausmann, R. G., Chi, M. T., & Roy, M. (2004). Proceedings of the annual meeting of the cognitive science mechanisms. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 26, pp. 547–552).

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*, 235–266.
Hosen, M., Ogbeibu, S., Giridharan, B., Cham, T.-H., Lim, W. M., & Paul, J. (2021). Individual motivation and social media influence on student knowledge sharing and learning performance: Evidence from an emerging economy. *Computers & Education, 172*, Article 104262. https://doi.org/10.1016/j.compedu.2021.104262. URL: https://www.sciencedirect.com/science/article/pii/S0360131521001391.
Huang, C.-m., Andrist, S., Sauppé, A., & Mutlu, B. (2015). *Using gaze patterns to predict task intent in collaboration* (Vol. 6, pp. 1–12). https://doi.org/10.3389/fpsyg.2015.01049
Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research, 81*, 11–24. https://doi.org/10.1016/j.ijer.2016.10.006. URL:, 10.1016/j.ijer.2016.10.006.
Ivtzan, I., Lomas, T., Wong, P., & Niemiec, R. (2015). *Second wave positive psychology: Embracing the dark side of life*.
Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing debates in learning and educational research: Toward a complex systems conceptual framework of learning. *Educational Psychologist, 51*, 210–218. https://doi.org/10.1080/00461520.2016.1166963
Järvelä, S., Gašević, D., Seppänen, T., Pechenizkiy, M., & Kirschner, P. A. (2020). Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *British Journal of Educational Technology, 51*, 2391–2406. https://doi.org/10.1111/bjet.12917
Jordan, B., Henderson, A., Jordan, B., & Henderson, A. (1995). Interaction analysis : Foundations and practice. *The Journal of the Learning Sciences, 4*, 39–103.
Jordan, M. E., & McDaniel, R. R., Jr. (2014). Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity. *The Journal of the Learning Sciences, 23*, 490–536. https://doi.org/10.1080/10508406.2014.896254
Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *I. J. Computer-Supported Collaborative Learning, 6*, 39–56. https://doi.org/10.1007/s11412-011-9109-9
Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*, 45–83. https://doi.org/10.1080/10508406.2011.591717. URL:, 10.1080/10508406.2011.591717.
Kapur, M., & Kinzer, C. K. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning, 4*, 21–46. https://doi.org/10.1007/s11412-008-9059-z. URL:, doi:10.1007/s11412-008-9059-z.
Käser, T., Klingler, S., Schwing, A. G., & Gross, M. (2017). Dynamic bayesian networks for student modeling. *IEEE Transactions on Learning Technologies, 10*, 450–462. https://doi.org/10.1109/TLT.2017.2689017
Kashdan, T., & Biswas-Diener, R. (2014). *The Upside of your dark side: Why being your whole self–not just your "good" self–drives Success and fulfillment*. Penguin Publishing Group. URL: https://books.google.ch/books?id=C5QxAwAAQBAJ.
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*, 187–215. https://doi.org/10.1007/s11409-014-9112-4. URL:, 10.1007/s11409-014-9112-4.
Kollar, I., Fischer, F., & Hesse, F. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*. https://doi.org/10.1007/s10648-006-9007-2
Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., & Lampi, E. (2021). What do we do when we analyse the temporal aspects of computer-supported collaborative learning? A systematic literature review. *Educational Research Review, 33*, Article 100387. https://doi.org/10.1016/j.edurev.2021.100387. URL: https://www.sciencedirect.com/science/article/pii/S1747938X21000105.
Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., & Mannonen, J. (2020). The potential of temporal analysis: Combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Computers & Education, 143*, Article 103674. https://doi.org/10.1016/j.compedu.2019.103674. URL: https://www.sciencedirect.com/science/article/pii/S0360131519302271.
Lim, W. M. (2020). A typology of student diversity and an inclusive student learning support system: Insights for higher education. *Educational Practice and Theory, 42*, 81–87. https://doi.org/10.7459/ept/42.1.06
Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*, 693–715. https://doi.org/10.1007/s10648-016-9379-x. URL:, doi:10.1007/s10648-016-9379-x.
Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior, 52*, 562–572. https://doi.org/10.1016/j.chb.2015.03.082. URL:, 10.1016/j.chb.2015.03.082.
Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Monés, A., Kay, J., & Yacef, K. (2013). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *International Journal of Computer-Supported Collaborative Learning, 8*, 455–485. https://doi.org/10.1007/s11412-013-9184-1. URL: http://link.springer.com/10.1007/s11412-013-9184-1.
Menon, D., Bp, S., Romero, M., & Viéville, T. (2019). Going beyond digital literacy to develop computational thinking in K-12 education. In L. Daniela (Ed.), *Smart pedagogy of digital learning*. Taylor&Francis (Routledge).
Nasir, J., Bruno, B., Chetouani, M., & Dillenbourg, P. (2021a). What if social robots look for productive engagement? *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-021-00766-w
Nasir, J., Bruno, B., & Dillenbourg, P. (2021b). PE-HRI-temporal: A Multimodal Temporal Dataset in a robot mediated Collaborative Educational Setting. URL: https://doi.org/10.5281/zenodo.5576058.
Nasir, J., Kothiyal, A., Bruno, B., et al. (2021c). Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. *International Journal of Computer-Supported Collaborative Learning, 16*, 485–523. https://doi.org/10.1007/s11412-021-09358-2.
Nasir, J., Norman, U., Bruno, B., & Dillenbourg, P. (2020). When positive perception of the robot has no effect on learning. In *In 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*.
Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology, 51*, 1527–1547. URL: https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12982. https://doi.org/10.1111/bjet.12982. arXiv:https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.12982.
Paans, C., Onan, E., Molenaar, I., Verhoeven, L., & Segers, E. (2019). How social challenges affect children's regulation and assignment quality in hypermedia: A process mining study. *Metacognition and Learning, 14*, 189–213. https://doi.org/10.1007/s11409-019-09204-9. URL:, 10.1007/s11409-019-09204-9.
Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering, 21*, 759–772. https://doi.org/10.1109/TKDE.2008.138
Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Proceedings of the 28th international conference on neural information processing systems (Volume 1* NIPS'15, pp. 505–513). Cambridge, MA, USA: MIT Press.
Ramachandran, A., Huang, C.-M., & Scassellati, B. (2019a). Toward effective robot–child tutoring: Internal motivation, behavioral intervention, and learning outcomes. *ACM Transactions on Interactive Intelligent Systems, 9*, 1–23. https://doi.org/10.1145/3213768
Ramachandran, A., Sebo, S. S., & Scassellati, B. (2019). *Personalized robot tutoring using the assistive tutor POMDP (AT-POMDP)* (pp. 1–8). Proceedings of The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI).
Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in cscl research. *I. J. Computer-Supported Collaborative Learning, 4*, 239–257. https://doi.org/10.1007/s11412-009-9070-z
Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: An analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction, 29*, 375–415.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer-supported collaborative learning* (pp. 69–97).
van de Sande, B. (2013). Properties of the bayesian knowledge tracing model. In *EDM 2013*.
Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 13*, 241–261.
Schulte, P. L. (1996). A definition of constructivism. *Science Scope, 20*, 25–27.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10*, 70–76.
Sharma, K., Papamitsiou, Z., Olsen, J. K., & Giannakos, M. (2020). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. In *Proceedings of the tenth international Conference on learning analytics & knowledge LAK '20* (pp. 480–489). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3375462.3375498. URL:, 10.1145/3375462.3375498.
Siegler, R. S., & Crowley, K. (1991). The microgenetic method : A direct means for studying cognitive development. *American Psychologist, 46*, 606–620. https://doi.org/10.1037/0003-066X.46.6.606
Sinha, T. (2021a). Enriching problem-solving followed by instruction with explanatory accounts of emotions. *The Journal of the Learning Sciences*, 1–48.
Sinha, T. (2021b). Enriching problem-solving followed by instruction with explanatory accounts of emotions. *The Journal of the Learning Sciences*, 1–48. https://doi.org/10.1080/10508406.2021.1964506, 00, 10.1080/10508406.2021.1964506.
Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works. *Evidence for Productive Failure, XX*. https://doi.org/10.3102/00346543211019105
Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning, 10*, 273–307.
Sobocinski, M., Malmberg, J., & Järvelä, S. (2017). Exploring temporal sequences of regulatory phases and associated interactions in low- and high-challenge collaborative learning sessions. *Metacognition and Learning, 12*, 275–294. https://doi.org/10.1007/s11409-016-9167-5
Spikol, D., Ruffaldi, E., & Cukurova, M. (2017). Using multimodal learning analytics to identify aspects of collaboration in project-based learning. *Computer-Supported Collaborative Learning Conference, CSCL, 1*, 263–270. https://doi.org/10.22318/cscl2017.37
Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, 34.
Viswanathan, S. A., & Vanlehn, K. (2018). Using the tablet gestures and speech of pairs of students to classify their collaboration. *IEEE Transactions on Learning Technologies, 11*, 230–242. https://doi.org/10.1109/TLT.2017.2704099

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*. https://doi.org/10.1007/s10648-016-9361-7

Voutsina, C., George, L., & Jones, K. (2019). Microgenetic analysis of young children's shifts of attention in arithmetic tasks: Underlying dynamics of change in phases of seemingly stable task performance. *Educational Studies in Mathematics, 102*, 47–74. https://doi.org/10.1007/s10649-019-09883-w. URL:, doi:10.1007/s10649-019-09883-w.

Wang, C., Sahebi, S., Zhao, S., Brusilovsky, P., & Moraes, L. O. (2021). Knowledge tracing for complex problem solving: Granular rank-based tensor factorization. In *Proceedings of the 29th ACM conference on user modeling, adaptation and personalization* (pp. 179–188). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3450613.3456831. URL:.

Yang, C. W., Cukurova, M., & Porayska-Pomsta, K. (2021). *Dyadic joint visual attention interaction in face-to-face collaborative problem-solving at K-12 maths education: A multimodal approach.* CEUR Workshop Proceedings.