



Detecting Collaborative Learning Through Emotions: An Investigation Using Facial Expression Recognition

Yugo Hayashi(✉)

College of Comprehensive Psychology, Ritsumeikan University,
2-150 Iwakura-cho, Osaka, Ibaraki 567-8570, Japan
y-hayashi@acm.org

Abstract. Providing adaptive feedback to learners engaging in collaborative learning activities is one research topic in the development of intelligent tutoring systems. However, there is a need to investigate how to systematically evaluate a learner's activities and provide feedback on them. The present study investigates how emotional states, detected through facial recognition, can be utilized to capture the learning process in a simple jigsaw-type collaborative task. It was predicted that when learners argue with each other and reason deeply, they may experience several emotional states such as positive and negative states. The results show that when learners work harder on developing a mutual understanding through conflictive interaction, negative emotions can be used to predict this process. This study contributes to the knowledge of how emotional states detected by facial recognition technology can be applied to predict learning process in conflictive tasks. Moreover, these empirical results will impact the development of adaptive feedback mechanisms for intelligent tutoring systems for collaborative learning.

Keywords: Collaborative learning · Pedagogical conversational agents · Emotion · Learning assessment

1 Introduction

Intelligent tutoring systems have been long investigated in educational science [3, 8, 9, 16, 21, 28], and one of the goals of these systems is to detect the learners' mental states and adaptively provide feedback. The use pedagogical conversational agents (PCAs) demonstrating benefits in learning gains has emerged in the last decade [8, 20, 22]. Recently, social learning such as learner-learner collaborative learning has come to be regarded as an important skill for the 21st century, and several studies have used PCAs in the context of collaborative learning. However, in cognitive and learning science, the mechanisms of collaborative interactions and their related process are not fully understood. This paper demonstrates how facial expression recognition can be used to predict emotional states to evaluate collaboration.

1.1 Collaborative Learning and Intelligent Systems

Studies in cognitive science show that collaboration helps to externalize knowledge [25,27] as well as facilitate meta-cognition during explanations [4] and perspective change [14]. It has been noted that social-conflict-based learning plays an important role in the learning process [19,29], and collaborative learning takes advantage of the nature of such conflict-based learning. Several studies in this area have attempted to understand these activities [7,15]. The 2015 Programme for International Student Assessment, which is administered by the Organisation for Economic Co-operation and Development, has surveyed student skills and knowledge [26]. In these surveys, two skills that leverage collaborative learning, task-work and team-work, are considered to be important skills. The former is related to problem-solving skills and the latter is related to social skills such as the ability to coordinate and establish common ground with other group members.

Although team-work plays an important role in collaborative learning, it has been difficult to quantitatively evaluate a learner's conversational behaviors with respect to the success or failure of team interactions. It is hence a challenge to develop computational models of a learner's interactions to automatically detect his/her state and provide group facilitation accordingly. There have been studies that have successfully detected a learner's state from linguistic data and used a PCA to assist learning [18]. However, it is still difficult to completely understand the detailed context of social interactions. There have also been attempts to use multiple variables such as verbal and nonverbal channels [5]; however, it is not fully understood which paradigm is best for evaluating both team-work and task-work in collaborative learning.

1.2 Using Emotional States as Predictors for Learning

Studies in collaborative learning have used tasks that generate social conflict, such as a jigsaw-type task, in which learners tend to discuss their different perspectives and conflictive states are expected to occur during the task [1]. On such occasions, it is likely that confusion and arguments may occur. As a result, learners may experience emotional states [6]. The study [8] showed that learning gains were positively correlated with confusion and engagement/flow, negatively correlated with boredom, and were uncorrelated with the other emotions. Moreover, psychology studies on general problem solving have discovered that when problem solvers are confronted in an impasse, the emotional states are highly related [24]. According to these studies, positive emotions are highly related to aptitude tasks. Other studies have also shown that positive emotions play an important role in interactive behavior [10].

These studies imply that emotional states, especially emotional states that are related to negative/positive feelings can be used to detecting a learner's performance and role in collaborative learning in a conflictive task. However, it is important to consider that the instances of confusion that are peripheral to the learning activity are unlikely to have any meaningful impact on learning [6].

In the present study, we use a jigsaw-type task that includes the integration of other learners' different perspectives. It is expected that, to establish common ground in order to achieve the task, learners will experience emotional states: a negative state during confusion and conflicts during and a positive state when communication is successful.

1.3 Goal and Hypotheses

The goal of this study is to investigate how an emotional state that is detected during collaborative learning can be used to predict performance in a conflictive collaborative learning task. The long-term goal of this research is to develop an adaptive collaborative tutoring system in which PCAs (developed in the authors' previous work) facilitate learning according to the learners' states.

In this study, we investigate collaborative learning in a jigsaw-type task in which socio-cognitive conflict is expected to occur. It is predicted that, to achieve the task, learners may experience both positive and negative emotions due to the nature of the task. We hence consider the following hypothesis (H1): when experiencing arguments during the task, learners become conflictive and confused, and these can be detected as negative emotions. Hypothesis H1 has two parts: more strongly negative emotions are related to higher-level coordination activities such as establishing common ground about their different knowledges (H1-a) and thus affect learning performance (H1-b). We also consider the following hypothesis (H2): learners experience positive emotions when establishing common ground and reaching agreement (H2-a) and these emotions thus influence learning performance (H2-b).

The present study investigates these hypotheses by focusing on emotions detected using learners' facial expressions. This use of artificial intelligence technology supports our long-term goal of developing intelligent and adaptive tutoring systems.

2 Method

2.1 Procedure and Task

Twenty Japanese university students participated in dyads in this experiment. The participants received course credit for participation. This study was conducted after passing an ethical review conducted by the authors' institutional ethical review committee.

The experiment design consisted of a pre-test, main task, and post-test procedure. The main task of this experiment was to explain a topic that was taught in a cognitive science class. They were required to explain the phenomenon of how humans process language information and were required to use two sub-concepts: "top-down processing" and "bottom up processing." This study adopts the jigsaw method [1], which is a style of learning in which each learner has knowledge of one of the sub-concepts and exchanges it with their

partner through explanation. The learners’ goal was to explain their different perspectives and provide an overall integrated explanation of the phenomenon using the two sub-concepts. To achieve their goal, they were required to argue about how each sub-concept can be used to explain the main concept.

Participants individually worked on the pre-test to determine whether they already knew about the sub-concepts. The main task was conducted for ten minutes. After completion, the learners again individually performed the post-test so that their knowledge gain could be measured.

2.2 Experimental Set-Up

The experiment was conducted in a designated laboratory experiment room. A redeveloped version of the system designed in a previous study was used [10–12]. Learners sat in front of a computer display and talked to each other orally. The experimental system was developed in the Java language and run on an in-house server-client network platform. The two learners’ computers were connected through a local area network, and task execution was controlled by a program on the server. The system also features a conversational PCA that provided meta-cognitive suggestions [10] to facilitate their explanations. The example of the displays are shown in Fig. 1.

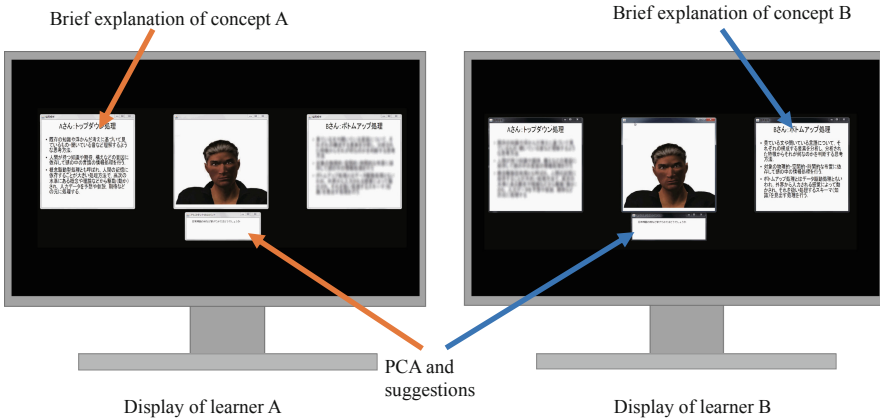


Fig. 1. Example of participants’ screens.

An embodied PCA was presented in the center of the screen, which physically moved when it spoke. Below the PCA, there was a text box that showed messages. The experimenter sat to the side in the experiment room and manually instructed the PCA to provide meta-cognitive suggestions. The suggestions were made once per minute when there was a gap in conversation. Five types of meta-cognitive suggestions were used, such as reminding learners to achieve the task goal [2] and facilitating metacognition [11].

During the task, the experimenter also video-recorded the learner's facial expressions and recorded their voice during the main task. The facial recordings were used as one measure to understand the learner's affective state during their task, as explained further in the next section. All of the recorded conversations were transcribed into text to further analyze the quality of the explanations.

2.3 Measures

This study examined the performance of the participants on the pre- and post-tests and the quality of the collaboration through coding the learners' performance.

Pre-and Post-tests. The responses to the test were coded as follows: 0 = a wrong explanation, 1 = a naive but correct explanation, 2 = a concrete explanation based on the presented materials, and 3 = a concrete explanation based on the presented materials that used examples and metacognitive interpretations. Two coders conducted this analysis with an accuracy of 78%. They discussed any mismatching codes to determine the final codes.

Quality of Collaboration. The study adopted part of the coding scheme from [23] that are related to emotion capture. The original full scheme is as follows: 1 = mutual understanding, 2 = dialogue management, 3 = information pooling, 4 = reaching consensus, 5 = task division, 6 = time management, 7 = reciprocal interaction, and 8 = individual task orientation. The present study excluded the codes 3, 5, 6, and 7 because they are inappropriate for this study. Just as for the pre- and post-tests, two coders evaluated the transcripts of the experiment dialogues, with a coding match of 85%.

Facial Expressions. For the facial expression analysis, this study used Face Reader (<https://www.noldus.com/>) to evaluate the emotional states of the learners during the interactions. This system can classify expressions as one of the emotional categories of joy, anger, sadness, surprise, fear, or disgust [17]. The tool recognizes fine-grained facial features based on minimal muscular movements described by the Facial Action Coding System (FACS). The systems use Active Appearance Model (AAM) for creating a model of the facial expressions for classification, where the shape of the face is defined by a shape vector that contains the coordinates of the landmark points (for details of the algorithm, see [17]). In addition, the two coders checked the reliability of the automated coding by randomly selecting, manually coding, and checking the accuracy of the automatically detected emotional states. The accuracy of the recognition system was 72%. In this study, we calculated the variable of each emotional state for each individual and used it as a representative value for analysis. Using these values as predictors, we investigated how they can predict the learning performance and collaboration process.

3 Results

3.1 Performance on the Pre- and Post-tests

For each individual learner, the gain in score between the pre- and post-tests was calculated ($\text{gain} = [\text{post-test score}] - [\text{pre-test score}]$). Pearson’s correlation coefficients were calculated to determine if there was correlation between the major detected states and the gain. Table 1 shows the results, which show that no correlation was detected between emotional state and performance gain.

Table 1. Results of correlation analysis for the performance and emotional states: “n.s.” = no significance.

Neutral	Happy	Sad	Angry	Surprised	Scared	Disgusted
-0.183 n.s.	0.245 n.s.	-0.045 n.s.	0.077 n.s.	-0.138 n.s.	0.1 n.s.	-0.056 n.s.

3.2 Collaboration Process

To investigate the relationships among the learning process and emotional states, we conducted a Pearson’s correlation analysis for the coded variables and each emotional state. Table 2 shows the results.

Table 2. Results of correlation analysis for the collaboration process and emotional states: “n.s.” = no significance, “+” = marginal significance ($p < .10$), and “*” = significance ($p < .05$).

	Neutral	Happy	Sad	Angry	Surprised	Scared	Disgusted
1. Mutual understanding	-0.278 +	0.194 n.s.	-0.071 n.s.	0.302 *	0.056 n.s.	-0.226 +	0.088 n.s.
2. Dialogue management	-0.01 n.s.	-0.249 +	-0.177 n.s.	0.285 +	0.253 +	0.046 n.s.	-0.003 n.s.
4. Reaching consensus	-0.202 n.s.	0.340 *	-0.098 n.s.	0.301 *	-0.116 n.s.	-0.086 n.s.	0.021 n.s.
8. Individual task orientation	-0.4 *	0.348 *	-0.077 n.s.	0.283 n.s.	-0.052 n.s.	-0.065 n.s.	0.236 n.s.

The following sections further analyze the regressions for each type of collaborative process.

“Mutual Understanding”. The Pearson’s correlation analysis shows that there was a significant correlation between “mutual understanding” and “angry.” As predicted, this could be due to the fact that learners working hard to develop mutual understandings in this jigsaw-like task, learners experienced more interpersonal conflict and expressed angry facial expressions. A multiple regression analysis was performed in which learning gain was regressed for each of the variables. The regression coefficient R^2 was .385 and the analysis of variance (ANOVA) F -value was 2.322, indicating significance for both variables ($p = .05$.) The results suggest that the degree to which the process of trying to establish common ground can be predicted by facial expressions displaying anger. This supports our hypothesis H1-a.

“Dialogue Management”. The results of the Pearson’s correlation analysis show that there were no significant correlations among any of the variables for “dialogue management.” However, some marginal effects were found for the “happy,” “angry,” and “surprised” emotions.

“Reaching Consensus”. The results of the Pearson’s correlation analysis show that there were significant correlations between “reaching consensus” and “happy” and “angry” emotions. Learners may have experienced happiness because they had reached consensus. Anger may have appeared because learners experienced conflicts about their different perspectives and/or frustration prior to reaching consensus. To further investigate the prediction ability of these emotions, a multiple regression analysis was conducted. However, the regression coefficient R^2 was .316 and the ANOVA F -value was 1.717, indicating no significance for both variables ($p = .14$.)

“Individual Task Orientation”. The results of the Pearson’s correlation analysis show that there were significant correlations between “individual task orientation” and “happy” and “neutral.” This indicates that when learners worked well individually they were engaging with the task with positivity. To further investigate the predictability, we conducted a multiple regression analysis where learning performance (the dependent variable) was regressed on each of the variables. The regression coefficient R^2 was .358 and the ANOVA F -value was 2.072, indicating only marginal significance ($p = .08$.)

4 Discussion and Conclusions

This study focused on learning situations in which learners were engaged in a jigsaw-based collaborative learning task, which requires interactive conflicts to achieve the goal. Using this task, the author investigated emotional states, which have been recently employed as important indicators for understanding learners’ internal processes. Facial recognition technology was used to estimate the learners’ facial expressions, which were then used to reveal the relationships between the social interactive process and learning performance while using a tutoring system. To investigate this, this study used a collaborative learning platform designed by the authors in which an embodied PCA was embedded. This study hypothesized that both positive and negative emotional states can capture the process of several types of interaction process such as developing common ground and furthermore learning performance. However, the results show that emotional states were useful to predict only for the learning process. More specifically, negative emotions detected from learners’ facial expressions were able to predict the process of developing mutual understandings (supporting hypotheses H1-a.) This is considered that when developing consensus in this task, learners have to integrate their different perspectives and thus require interpersonal conflicts, which may be associated with confusions which can be detected as negative emotions.

Further investigation should be conducted by directly examining the degree of confusions such as dependent variables used by [6], for future work. Moreover, combinations of using several different variables should provide a larger view of the relationships between the interaction process and the types of emotional states of the learners.

On the other hand, none of the detected emotions were useful for predicting learning performance (not supporting hypotheses H1-b, and H2-b.) For this point, it can be discussed that gaining knowledge in this task was more-like an individual activity and there was hence no need for a learner to express his/her emotional state through facial expressions when thinking and reasoning. Some studies point that collaborative problem-solving is composed by phases of (1) task work which is to build internal knowledge and (2) team work (coordination) which is to exchange and share internalized knowledge to build collective knowledge [7]. It can be considered that each factors should be supported individually by using different types of facilitation methods [13]. Also, our hypotheses H1-b and H2-b might hold if we further investigate on other emotional or affective states using different measures.

In conclusion, these results will contribute to development of intelligent tutoring systems because the results show that learners' interaction process can be detected from emotional states. On designing such systems, detecting individuals' emotional states accurately is one of the challenges in artificial intelligence, but recently there has been many successes in the software development and this study is one good example. Past studies in ITS has not yet fully demonstrated how facial expressions can be automatically collected and used for detecting their emotional states especially in a task in a conflict-based collaborative learning environment. This paper contributes to ITS by showing empirical results based on laboratory study, showing how such technology enables to detect emotional states of learners conversational process. The next aim of this research is to develop a classifier based on our results and then develop a system that provides adaptive PCA facilitation based on the real-time recognition of the learner's states.

Acknowledgments. This work was supported by Grant-in-Aid for Scientific Research (C), Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 16K00219.

References

1. Aronson, E., Patnoe, S.: *The Jigsaw Classroom: Building Cooperation in the Classroom*, 2nd edn. Addison Wesley Longman, New York (1997)
2. Azevedo, R., Cromley, J.: Does training on self-regulated learning facilitate students' learning with hypermedia? *J. Educ. Psychol.* **96**(3), 523–535 (2004)
3. Biswas, G., Leelawong, K., Schwartz, D., Vye, N.: Learning by teaching: a new paradigm for educational software. *Appl. Artif. Intell.* **19**(3), 363–392 (2005)
4. Chi, M., Leeuw, N., Chiu, M., Lavancher, C.: Eliciting self-explanations improves understanding. *Cogn. Sci.* **18**(3), 439–477 (1994)

5. von Davier, A.A., Hao, J., Liu, L., Kyllonen, P.: Interdisciplinary research agenda in support of assessment of collaborative problem solving: lessons learned from developing a collaborative science assessment prototype. *Comput. Hum. Behav.* **76**, 631–640 (2017)
6. D’Mello, S., Lehman, B., Pekrun, R., Graesser, A.: Confusion can be beneficial for learning. *Learn. Instr.* **29**, 153–170 (2014)
7. Fiore, S.M., Rosen, M.A., Smith-Jentsch, K.A., Salas, E., Letsky, M., Warner, N.: Toward an understanding of macrocognition in teams: predicting processes in complex collaborative contexts. *Hum. Factors* **52**, 203–224 (2010)
8. Graesser, A., Chipman, P., Haynes, B., Olney, A.: AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Trans. Educ.* **48**(4), 612–618 (2005)
9. Graesser, A., McNamara, D.S.: Computational analyses of multilevel discourse comprehension. *Top. Cogn. Sci.* **3**(2), 371–398 (2011). <https://doi.org/10.1111/j.1756-8765.2010.01081.x>
10. Hayashi, Y.: On pedagogical effects of learner-support agents in collaborative interaction. In: *Proceeding of the 11th International Conference on Intelligent Tutoring Systems (ITS 2012)*, pp. 22–32 (2012)
11. Hayashi, Y.: Togetherness: Multiple pedagogical conversational agents as companions in collaborative learning. In: *Proceeding of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014)*, pp. 114–123 (2014)
12. Hayashi, Y.: Coordinating knowledge integration with pedagogical agents: effects of agent gaze gestures and dyad synchronization. In: *Proceeding of the 13th International Conference on Intelligent Tutoring Systems (ITS 2016)*, pp. 254–259 (2016)
13. Hayashi, Y.: Gaze feedback and pedagogical suggestions in collaborative learning. In: Nkambou, R., Azevedo, R., Vassileva, J. (eds.) *ITS 2018*. LNCS, vol. 10858, pp. 78–87. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91464-0_8
14. Hayashi, Y.: The power of a “maverick” in collaborative problem solving: an experimental investigation of individual perspective-taking within a group. *Cogn. Sci.* **42**(S1), 69–104 (2018). <https://doi.org/10.1111/cogs.12587>
15. Hesse, F., Care, E., Buder, J., Sassenberg, K., Griffin, P.: A framework for teachable collaborative problem solving skills. In: Griffin, P., Care, E. (eds.) *Assessment and Teaching of 21st Century Skills*. EAIA, pp. 37–56. Springer, Dordrecht (2015). https://doi.org/10.1007/978-94-017-9395-7_2
16. Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M.A.: Intelligent tutoring goes to school in the big city. *Int. J. Artif. Intell. Educ. (IJAIED)* **8**, 30–43 (1997)
17. Kuilenburg, v.H., Wiering, M., Uyl, d.M.: A model based method for automatic facial expression recognition. In: *Proceedings of the European Conference on Machine Learning (ECML2005)*, pp. 194–205 (2005)
18. Kumar, R., Rose, C.: Architecture for building conversational architecture for building conversational agents that support collaborative learning. *IEEE Trans. Learn. Technol.* **4**(1), 21–34 (2011)
19. Lave, J., Wenger, E.: *Situated Learning - Legitimate Peripheral Participation*. Cambridge University Press, Cambridge (1991)
20. Leelawong, K., Biswas, G.: Designing learning by teaching agents: the Betty’s brain system. *Int. J. Artif. Intell. Educ.* **18**(3), 181–208 (2008)
21. Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G.J., Koedinger, K.R.: Studying the effect of a competitive game show in a learning by teaching environment. *Int. J. Artif. Intell. Educ.* **23**(1), 1–21 (2013)

22. McNamara, D., O'Reilly, T., Rowe, M.: *iSTART: A Web-Based Tutor that Teaches Self-explanation and Metacognitive Reading Strategies*. Lawrence Erlbaum Associates, Mahwah (2007)
23. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. *Int. J. Comput.-Support. Collab. Learn.* **2**(1), 63–86 (2007). <https://doi.org/10.1007/s11412-006-9005-x>
24. Metcalfe, J., Wiebe, D.: Intuition in insight and noninsight problem solving. *Mem. Cogn.* **15**(3), 238–246 (1987)
25. Miyake, N.: Constructive interaction and the interactive process of understanding. *Cogn. Sci.* **10**(2), 151–177 (1986)
26. OECD: *PISA 2015 Results (Volume V): Collaborative Problem Solving*. OECD Publishing, Paris (2017). <https://doi.org/10.1787/9789264285521-en>
27. Shirouzu, H., Miyake, N., Masukawa, H.: Cognitively active externalization for situated reflection. *Cogn. Sci.* **26**(4), 469–501 (2002)
28. VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rosé, C.P.: When are tutorial dialogues more effective than reading? *Cogn. Sci.* **31**(1), 3–62 (2007). <https://doi.org/10.1080/03640210709336984>
29. Vygotsky, L.S.: *The Development of Higher Psychological Processes*. Harvard University Press, Cambridge (1980)