Development and Evaluation of a Matching System to Facilitate Online Collaborative Learning

Haruka TADA[a*] & Fumihide TANAKA[b]
[a] *Graduate School of Science and Technology, University of Tsukuba, Japan*
[b] *Institute of Systems and Information Engineering, University of Tsukuba, Japan*
*haruka.tada@ftl.iit.tsukuba.ac.jp

Abstract: Collaborative learning in programming education is known to have educational benefits and is now practiced worldwide. However, the quality of collaborative learning varies depending on the extent to which teachers can support it. Currently, teachers are often too busy to provide sufficient support for collaborative learning. Therefore, in this study, we focused on one form of support, learner matching, and developed a matching system that uses emotion recognition results from learners' natural speech. We conducted comparative experiments with existing methods to clarify whether group formation based on the results of heterogeneity in emotion recognition is more effective than the existing method. The results suggest that the proposed matching system has the potential to bring positive effects to those learners with lower foundational skills. In addition, in supplementary surveys, the majority of participants reported that they found it easier to learn with their matched partners using the proposed method, thus supporting the hypothesis that the proposed method is more effective than the existing method.

Keywords: Collaborative Learning, CSCL, STEM Education, Programming Education, e-Learning

1. Introduction

The importance of science, technology, engineering, and mathematics (STEM) education is increasing, and programming education is being promoted actively around the world (Sáez-López et al., 2016). Programming education has been shown to improve students' cognitive skills, such as creativity, reasoning, and mathematical skills, and collaborative learning is recommended in such education (Scherer et al., 2019). Collaborative learning is a teaching method in which students at various levels cooperate in small groups toward a common goal. Since they are responsible not only for their own learning but also for each other's learning, the success of one student can then help others succeed (Gokhale, 1995). Collaborative learning has educational benefits over individual learning, and students who learn through collaborative learning are known to be able to engage in learning sustainably and persistently (SAHIN et al., 2014).

However, collaborative learning is not always effective, and its effectiveness depends on the richness and intensity of the interactions between group members during the collaboration (Dillenbourg & Hong, 2008). The presence of a teacher is also necessary to shape the interactions within the group (Muirhead, 2000). However, given the current situation where educational disparities are widening worldwide, it can be difficult for teachers to provide sufficient support. In Japan, educational disparities have arisen due to teachers' increased workload. In Japan, programming education became mandatory in elementary schools in 2020 to foster programming thinking skills. However, in a survey of elementary school teachers, most of the respondents felt overwhelmed due to overwork, and 67 out of 79 respondents felt anxious about teaching programming education (Oshima et al., 2020), Therefore, in such a situation, it is difficult to provide programming education equally.

In the field of computer-supported collaborative learning (CSCL), several approaches support group interaction with computers. One of these approaches is group formation (Dillenbourg & Hong, 2008). Although group formation is important in collaborative learning, it is often overlooked, but proposing an appropriate group formation may make it possible to create beneficial groups (Muehlenbrock, 2006). Recent studies have used survey or test results, such as age, ability, and gender differences as criteria for group formation (Topping, 2010; Uchida, 2016). However, this is not necessarily related to the interactions between the group members, and it is necessary to consider heterogeneity regarding learners' characteristics, such as their personality and social attitudes (Kinjo, 2018).

Therefore, we assume that the heterogeneity of the results of emotion recognition analyzed from the natural utterances of learners is relevant for group interactions. In this study, we targeted the scenario of pair work in robot programming and developed a system of matching by the heterogeneity of emotion recognition results that automates group formation in collaborative learning. In addition, as a research question, we aimed to clarify whether group formation based on the heterogeneity of emotion recognition results is more effective than current conventional methods. The system was evaluated through experiments with college students, who are easily accessible for feedback.

2. Overview of a Matching System

We developed a matching system using emotion recognition results from natural speech data acquired from pre-lesson interactions as an indicator. To procure natural speech from learners, they will learn with a virtual partner in a pre-lesson as shown in Figure 1. The virtual partner is an avatar programmed in advance for speech that talks to the learner like a real child, saying things like "What should we do here?" or "Have you finished yet?" to match the mission of the pre-lesson. By recording conversations between the virtual partner and the learner, natural speech data from the learner's speech can be collected.

The reason for using a virtual partner rather than an actual person as a pre-lesson partner is to prevent changes in the learner's speech content in response to the characteristics of the partner. By using a virtual partner, speech data can be collected equally under the same conditions for everyone. These data are then analyzed, and a matching algorithm is used based on the analysis results to match learners with each other. The matching algorithm was created through the experiments in Chapter 3.

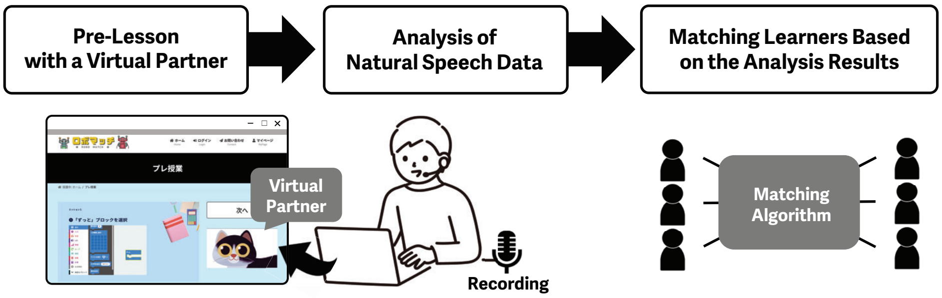

Figure 1. Matching Learners Based on the Analysis of Data from the Pre-Lesson.

3. Development of a Matching System

The experiment involved a total of 12 participants (6 males and 6 females) who were university undergraduate and graduate students. In all subsequent experiments in this study, individual names were not identifiable to by third parties, participation was voluntary, and there was no penalty for refusing to participate. Participants were informed of the purpose and content of this study and provided their verbal and written consent. This study was approved by the Ethics Committee of the University of Tsukuba (2022R676).

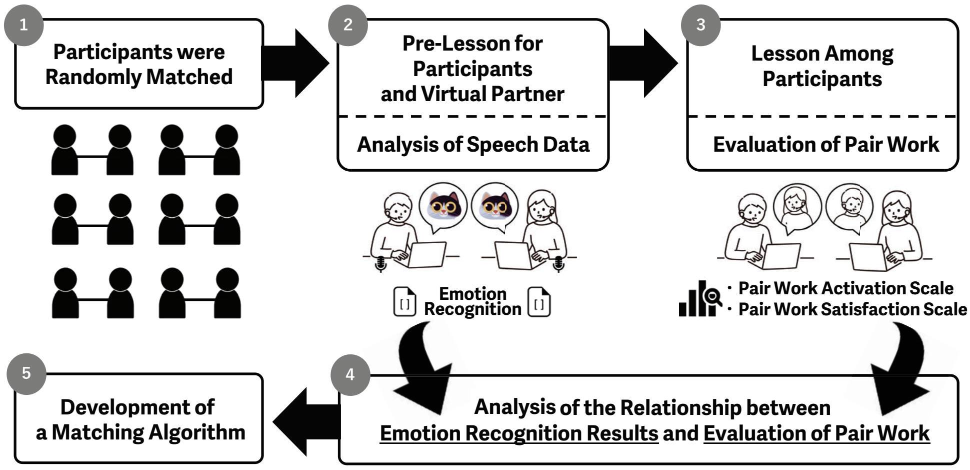

Figure 2. Flowchart of the Development of a Matching Algorithm.

3.1 Participants Were Randomly Matched

We randomly matched 12 participants and had them take a lesson in robot programming. The learning material for robot programming used in the experiment involved the use of micro:bit (micro:bit, 2023). Micro:bit is a small computer created by the British public broadcaster BBC to teach programming to children. As this was an online experiment, participants programmed the micro:bit using a simulator.

3.2 The Pre-Lesson for Participants and Virtual Partner/Analysis of Speech Data

Each participant was paired with a virtual partner for learning, and their speech data were collected during this process. The data were analyzed through emotion recognition as shown in Figure 3. First, those parts of the speech data that exceeded a certain amplitude for more than 0.6 seconds were identified as speech, and these segments of audio were extracted. Emotion recognition was performed using these speech data by inputting it into an emotion prediction model. The emotion prediction model was created using deep learning and the Ryerson Audio–Visual Database of Emotional Speech and Song (Livingstone & Russo, 2018), which is a dataset of 1440 audio samples to represent eight emotional states (neutral, calm, happy, sad, angry, fearful, disgusted, and surprised) stored in the data repository, Zenodo. The model structure uses Long Short-Term Memory, and the Mel-frequency cepstral coefficients (MFCC) were used as the feature. The reason for using MFCC as the feature is that it is effective for emotion recognition (Sato & Obuchi, 2007). The model was trained using training data, and 32 rounds of machine learning were performed to create an emotion prediction model with an accuracy of about 60% on test data.

Using this emotion prediction model, emotion recognition was performed. When the speech data of participant p's i-th speech was inputted into the emotion prediction model, an array x_p such as equation (1) was outputted. Here, $a_p(i)$ represents the neutral emotional level, $b_p(i)$ represents the calm emotional level, $c_p(i)$ represents the happy emotional level, $d_p(i)$ represents the sad emotional level, $e_p(i)$ represents the angry emotional level, $f_p(i)$ represents the fearful emotional level, $g_p(i)$ represents the disgusted emotional level, and $h_p(i)$ represents the surprised emotional level. For example, if $x_p(i)$ = [0.9, 0, 0.02, 0.04, 0, 0.04, 0, 0], it can be understood that the level of neutral emotion was high in the inputted speech data. These arrays were then averaged to create a single array. If the number of extracted speech data segments from participant p is n_p, then the averaged array X_p can be obtained as shown in equation (2). This X_p was used as the emotion recognition results for participant p.

$$x_p(i) = [a_p(i), b_p(i), c_p(i), d_p(i), e_p(i), f_p(i), g_p(i), h_p(i)] \tag{1}$$

$$X_p(i) = \left[\frac{1}{n_p}\sum_{i=1}^{n_p} a_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} b_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} c_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} d_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} e_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} f_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} g_p(i), \frac{1}{n_p}\sum_{i=1}^{n_p} h_p(i)\right] = [A_p, B_p, C_p, D_p, E_p, F_p, G_p, H_p] \tag{2}$$

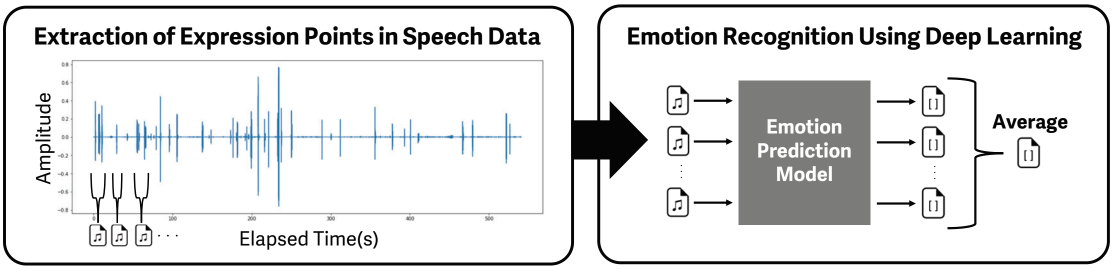

Figure 3. Analysis through Emotion Recognition Using Deep Learning.

3.3 Lesson Between Participants/Evaluation of Pair Work

Participants were paired up to engage in learning activities, and evaluations of their pair work were conducted. In assessing pair work, Uchida used activation and satisfaction as variables to verify the effectiveness of collaborative learning (Uchida, 2016). We use these two scales.

Activation was defined as the number of speech units, defined as a cohesive segment of speech, and counted from text data obtained by transcribing recorded conversations during pair work. Uchida used activation as an evaluation metric for cooperative learning, as it is believed that the number of utterances counted is effective in estimating the state of cooperative work and provides useful information for assessing judgments of pairs with poor cooperative relationships and interactions between group members. We divided the length of the class period into seconds because different pairs had different study times. The activation for participant a and participant b was denoted by $y_{a,b}$. In this experiment, pairs were matched as follows: (participant 0 and participant 1), (participant 2 and participant 3), (participant 4 and participant 5), (participant 6 and participant 7), (participant 8 and participant 9), and (participant 10 and participant 11), and the resulting data set of activation was $\{y_{0,1}, y_{2,3}, y_{4,5}, y_{6,7}, y_{8,9}, y_{10,11}\}$.

Table 1. *Satisfaction Questionnaire Evaluation Items (UCHIDA et al., 2016)*

1	Pair work was fun.
2	Pair work consultations were helpful.
3	I think pair work is a good method.
4	I think it is easier to answer in pairs than individually.
5	Pair work helped me to deepen my understanding.

Satisfaction was measured by a post-work survey that totaled the scores of each question item to determine the value of satisfaction. The survey questions are listed in Table 1, and in this experiment, participants answered the five questions in Table 1 on a 7-point scale (ranging from 1 to 7), and the sum of the scores for the pair of participants indicated satisfaction. Due to the influence of students' characteristics such as effort and motivation on course satisfaction, Uchida used satisfaction as an evaluation metric for cooperative learning, as it is believed to be useful in demonstrating the relationship between the pair effect and personality traits. The satisfaction for participants *a* and *b* was denoted by $z_{a,b}$. When the pairs were matched as (participant 0 and participant 1), (participant 2 and participant 3), (participant 4 and participant 5), (participant 6 and participant 7), (participant 8 and participant 9), and (participant 10 and participant 11), the resulting data set of satisfaction was $\{z_{0,1}, z_{2,3}, z_{4,5}, z_{6,7}, z_{8,9}, z_{10,11}\}$.

3.4 An Analysis of the Relationship Between Emotion Recognition Results and Evaluation of Pair Work

Let X_p be the array of emotional levels of participant p as represented in equation (2). When matching participants 0 through 11 as pairs (participant 0 and participant 1), (participant 2 and participant 3), (participant 4 and participant 5), (participant 6 and participant 7), (participant 8 and participant 9), and (participant 10 and participant 11), let $\{|A_0 - A_1|, |A_2 - A_3|, |A_4 - A_5|, |A_6 - A_7|, |A_8 - A_9|, |A_{10} - A_{11}|\}$ be the dataset of differences in neutral emotional levels between pairs, and let $\{A_0 + A_1, A_2 + A_3, A_4 + A_5, A_6 + A_7, A_8 + A_9, A_{10} + A_{11}\}$ be the dataset of the sums of neutral emotional levels between pairs. Similarly, define datasets for the differences and sums of seven other emotional states (calm, happy, sad, angry, fearful, disgusted, surprised), as well as for the differences and sums of the eight emotion recognition results, and for the datasets of activation and satisfaction (giving 32 combinations in total).

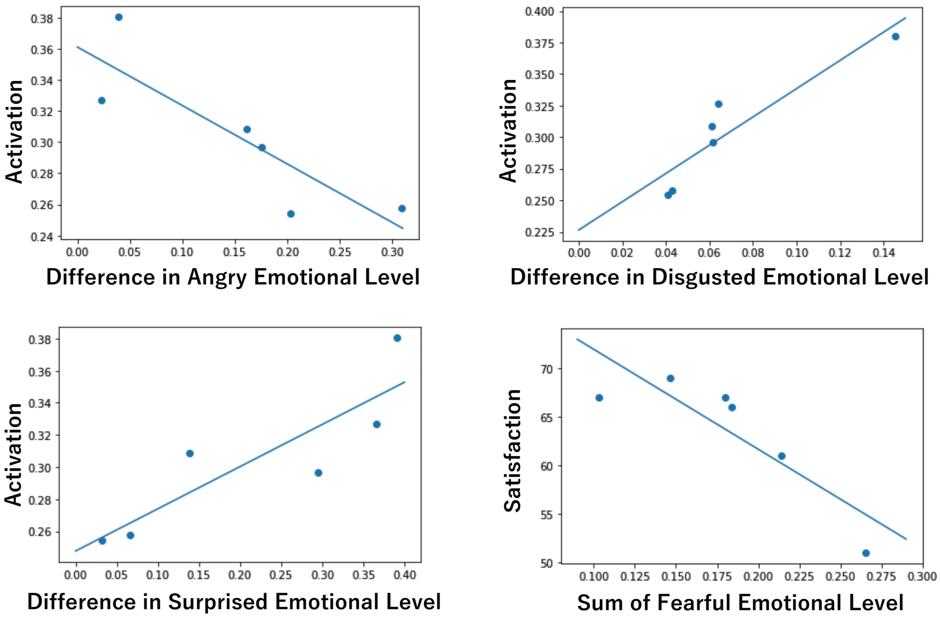

Figure 4. Relationship between Activations and Differences in Angry Emotional Levels (upper left), Disgusted Emotional Levels (upper right), and Surprised Emotional Levels (lower left). Relationship between Satisfaction and Sum of Fearful Emotional Levels (lower right).

In the analysis, we conducted a test of non-correlation, with the null hypothesis that "there is no correlation between the two variables" and the alternative hypothesis that "there is a difference between the population mean and the sample mean." We set the significance level to 0.05 and compared test statistic t, calculated from the correlation coefficient and a t-distribution value of $k = 2.7$. For all 32 combinations, we calculated test statistic t and found that the null hypothesis was rejected in four cases: the relationship between the dataset of differences in angry emotional levels and the dataset of activation ($t = 3.3 > k$), the relationship between the dataset of differences in disgusted emotional levels and the dataset of activation ($t = 4.7 > k$), the relationship between the dataset of differences in surprised emotional levels and the dataset of activation ($t = 3.5 > k$), and the relationship between the dataset of the sums of fearful emotional levels and the dataset of satisfaction ($t = 3.4 > k$). Therefore, we can say that the correlation coefficients between the dataset of differences in angry emotional levels and the dataset of activation (−0.86), between the dataset of differences in disgusted emotional levels and the dataset of activation (0.92), between the dataset of differences in surprised emotional levels and the dataset of activation (0.87), and between the dataset of sums of fearful emotional levels and the dataset of satisfaction, (−0.86) are significant at the 0.05 level. Graphs showing each relationship and the regression line created by the least squares method are shown in Figure 4. Assuming that the datasets for the differences in angry, disgusted, surprised, and the sum of fearful emotional levels are independent, we can consider the following four points in our discussion.

- The smaller the difference in the angry emotional level, the greater the activation.
- The larger the difference in the disgusted emotional level, the greater the activation.
- The larger the difference in the surprised emotional level, the greater the activation.
- The smaller the sum of the fearful emotional level, the greater the satisfaction.

3.5 The Development of a Matching Algorithm

After considering four aspects related to emotional levels, we developed a matching algorithm. The procedure for developing the algorithm is explained as follows:

First, we determined pairs of two individuals ($P1$, $P2$, $P3$, $P4$) from all the learners who wanted to be matched. $P1$ indicates the point that improves as the difference in the angry emotional level becomes smaller, $P2$ indicates the point that improves as the difference in the disgusted emotional level becomes larger, $P3$ indicates the point that improves as the difference in the surprised emotional level becomes larger, and P4 indicates the point that improves as the sum of the fearful emotional level becomes smaller. If we assume that the array of emotional levels for participant p is expressed as X_p, then the combinations of $P1$, $P2$, $P3$, and $P4$ for learners a and b are defined as expressions (3), (4), (5), and (6), respectively.

Next, assuming that the data groups for the difference in the angry emotional level, the difference in the disgusted emotional level, the difference in the surprised emotional level, and the sum of the fearful emotional level were all independent, we normalized these four data groups to have a minimum of 0 and a maximum of 1 to treat these four points equally. When $P1_{a,b}$, $P2_{a,b}$, $P3_{a,b}$, and $P4_{a,b}$ were denoted as the normalized data, $P1'_{a,b}$, $P2'_{a,b}$, $P3'_{a,b}$, and $P4'_{a,b}$, respectively, the overall point P for learners a and b was defined as expression (7).

Finally, as we contend that higher P values correspond to increased pair work activity and satisfaction, we recommend learners as a match in the order of their P values. For example, if we calculate the points such that $P_{0,1} < P_{0,2} < P_{0,3}$ among learners 0, 1, 2, and 3, we recommend learners 3, 2, and 1 in that order to learner 0.

$$P1_{a,b} = \left|\frac{1}{E_a - E_b}\right| \tag{3}$$

$$P2_{a,b} = |G_a - G_b| \tag{4}$$

$$P3_{a,b} = |H_a - H_b| \tag{5}$$

$$P4_{a,b} = \left|\frac{1}{F_a + F_b}\right| \tag{6}$$

$$P_{a,b} = P1'_{a,b} + P2'_{a,b} + P3'_{a,b} + P4'_{a,b} \tag{7}$$

4. Evaluation of a Matching System

An evaluation experiment was conducted on the proposed matching system. The participants consisted of a total of 10 university and graduate students (5 male and 5 female). The proposed method was evaluated by comparing it with a pairing method that used sex and basic academic skill, which has been confirmed to be effective by Uchida (Uchida, 2016). Uchida explains the pairing method using sex and basic academic skill as follows:
(1) Sort the students in descending order based on their basic academic skill scores.
(2) Pair them according to the sorting results.
(3) If the pairing results are in same-sex pairs in (2), replace them with opposite-sex pairs with a small difference in basic academic skill scores. However, if the difference in basic academic skill scores exceeds 2 out of 20 points, they will be kept as same-sex pairs.

An experiment was conducted to compare the existing method of matching using characteristics obtained from surveys or tests as indicators with the proposed method of matching using the heterogeneity related to learners' characteristics as indicators.

4.1 Experimental Methods

First, the experimental participants learned with partners matched by using the existing method. In Uchida's experiment, a 20-point test was conducted to measure basic academic skills. In this experiment, a programming test with a maximum score of 20 points was conducted to measure basic programming skills in robot programming. Then, based on Uchida's algorithm, 10 participants were matched and engaged in pair learning, after which pair work was evaluated using Uchida's activation and satisfaction.

Next, the participants learned with partners matched using the proposed method. All participants learned with virtual partners, and the speech data obtained were analyzed to match participants based on the matching algorithm developed. Ten participants were matched and engaged in pair learning, after which pair work was evaluated in the same manner. Finally, the pair work evaluation of the existing method and the proposed method were compared.

4.2 Experimental Results and Discussion

4.2.1 Comparison of the mean values of pair work evaluation

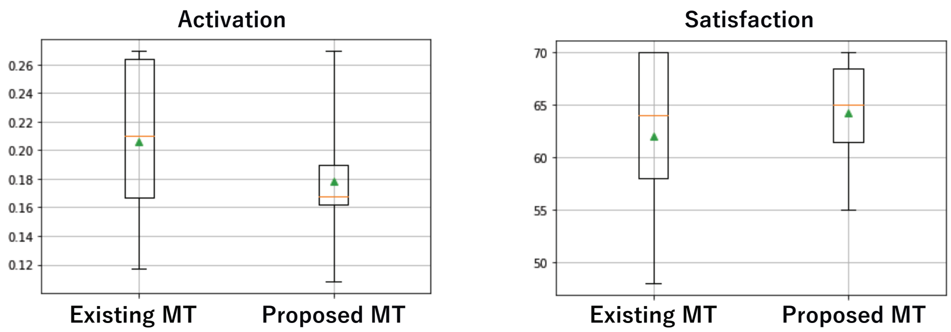

Figure 5. Comparison of Existing and Proposed Methods (MT) for Activation (left) and Satisfaction (right).

Figure 5 compares the activation and satisfaction of the existing method and the proposed method. The mean value of activation for the existing method was 0.21 and for the proposed method it was 0.18. Assuming that the pairs in the existing and proposed methods were independent since the participants were the same but the combinations were different, we performed an analysis to determine whether there was a difference in the mean values of pair work evaluation for both methods. A two-sided t-test without correspondence was performed for the analysis. The null hypothesis was that "there is no difference in the mean values of pair work evaluation between the existing and proposed methods," and the alternative hypothesis was that "there is a difference in the mean values of pair work evaluation between the existing and proposed methods." We compared the t-value with the boundary value of $k = 2.2$ for a two-sided test with a significance level of 0.05. As a result, the null hypothesis was not rejected, and it was found that there was no significant difference in the mean values of pair work evaluation between the existing and proposed methods, both for activation $(t = 0.83, p = 0.42)$ and satisfaction $(t = -0.54, p = 0.60)$.

4.2.2 Comparison of the mean values of pair work evaluation

To investigate the effectiveness of the proposed method over existing approaches for specific types of participants, we analyzed individual satisfaction (on a scale of 35 points). First, we examined the participants' basic academic skills (based on pre-test results), their satisfaction when matched and learning with partners using the existing method, and their satisfaction when matched and learning with partners using the proposed method. The differences in satisfaction (calculated by subtracting the satisfaction obtained with the existing method from that of the proposed method) are presented in Table 2, where participants with lower basic academic skills showed higher satisfaction with the proposed method compared to the existing approach. Consequently, we proceeded with an analysis to explore the relationship between basic academic skill scores and differences in satisfaction.

In the analysis, we conducted a test of non-correlation, with the null hypothesis that "there is no correlation between the two variables" and the alternative hypothesis that "there is a difference between the population mean and the sample mean." We set the significance level to 0.05 and compared the test statistic t, calculated from the correlation coefficient and a t-distribution value of $k = 2.7$. As a result, since the null hypothesis was rejected $(t = 3.3 > k)$, we can assert that the correlation coefficient (−0.84) between academic skill scores and differences in satisfaction is significant in terms of t. The relationship and the regression line created by the least squares method are shown in Figure 6. From there, it can be contended that <u>the proposed matching system may have a greater potential to yield positive effects for learners with lower basic academic skills</u>.

Table 2. *Personal Comparison of Satisfaction with the Existing and Proposed Methods (MT).*

Participant	Academic Skill Score	Satisfaction (Existing MT)	Satisfaction (Proposed MT)	Difference in Satisfaction
0	20	29	27	−2
1	20	29	26	−3
2	18	34	35	+1
3	16	30	33	+3
4	16	35	35	0
5	16	35	35	0
6	14	35	35	0
7	14	35	35	0
8	14	28	34	+6
9	8	20	29	+9

4.2.3 Direct evaluation through questionnaires allocated to participants

In addition to the satisfaction survey regarding pair work, we conducted interviews after all the experiments to ask which learning partner the participants found easier to study with and the reasons for their preferences. During the interviews, we deliberately concealed whether the partners were matched using the existing method or the proposed method. The results of the survey are presented in Figure 7. Among the participants, two individuals were matched with the same partners using both the existing and proposed methods, so we did not obtain survey responses from them. However, 70% of the participants stated that they found it easier to learn with partners matched through the existing method.

Furthermore, in the interviews with participants who rated the proposed method more favorably, the majority reported that their learning partners in the proposed method were easier to talk to, making it simpler to ask questions and seek advice. However, some participants mentioned that during the proposed method period, as it was their second class together, they felt less resistance in interacting with their partners compared to the first class with the existing method, possibly because they were getting accustomed to the experimental setup.

These results suggest that the order of participation and the nature of the tasks may have introduced bias among the participants. To support the statement made in section *4.2.2*, that "the proposed matching system may have a greater potential to produce positive effects for learners with lower basic academic skills," further experiments are needed. These experiments should include increasing the number of participants and ensure no bias by assigning different groups of participants to learn with either the existing or the proposed method.

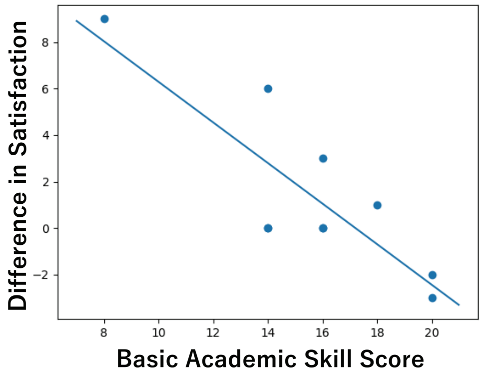

Figure 6. Relationship between Basic Academic Skill Scores and Differences in Satisfaction.

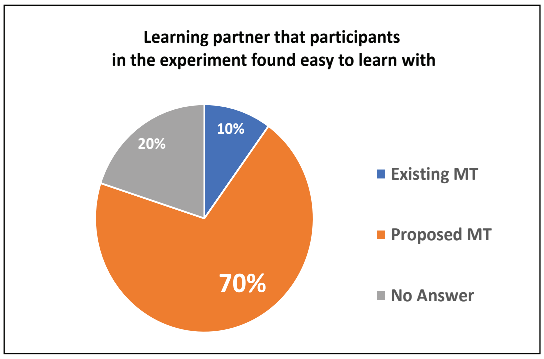

Figure 7. Survey results on "Learning partner that participants in the experiment found most easy to learn with."

5. Conclusion

In this study, we proposed a matching system for automating group formation by the heterogeneity of emotion recognition results in online collaborative learning during paired robot programming. We conducted an evaluation experiment comparing the existing and proposed methods of the matching system to clarify whether group formation based on the heterogeneity of emotion recognition results is more effective than using conventional methods. While no significant difference was observed in the average pairwise evaluation ratings, we found that the proposed method resulted in higher satisfaction ratings for those participants with lower basic skills. However, the experiment had limitations such as a small number of participants and thus insufficient data, and the possibility of bias due to the comparison of the existing and proposed methods with the same participants and different task content.

Therefore, in future studies, we plan to increase the number of participants and conduct sufficient testing to refine the matching algorithm. In addition, we plan to conduct between-participant comparisons with standardized task content, rather than within-participant comparisons. In this way, we will carefully account for biases due to differences in order of participation and task content, and continue with the evaluation experiments.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 19H01112.

References

Sáez-López, J.-M., Román-González, M., & Vázquez-Cano, E. (2016). Visual programming languages integrated across the curriculum in Elementary School: A two year case study using “scratch” in five schools. *Computers & Education*, *97*, 129–141. https://doi.org/10.1016/j.compedu.2016.03.003

Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of Learning Computer Programming: A meta-analysis of transfer effects. *Journal of Educational Psychology*, *111*(5), 764–792. https://doi.org/10.1037/edu0000314

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education, 7*(1). https://doi.org/10.21061/jte.v7i1.a.2

SAHIN, A., AYAR, M. C., & ADIGUZEL, T. (2014). Stem related after-school program activities and associated outcomes on student learning. *Educational Sciences: Theory & Practice, 14*(1), 309–322.

Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL Macro Scripts. *International Journal of Computer-Supported Collaborative Learning, 3*(1), 5–23. https://doi.org/10.1007/s11412-007-9033-1

Muirhead, B. (2000). Enhancing Social Interaction in Computer-Mediated Distance Education. *Journal of Educational Technology & Society, 3*(4), 1–11. http://www.jstor.org/stable/jeductechsoci.3.4.1

Oshima, T., Saito, H., & Okajima, Y. (2020). How Busy Teachers Feel About Upcoming Programing Education?. *Bulletin of Joetsu University of Education [published in Japanese], 40*(1), 33-43.

Muehlenbrock, M. (2006). Learning Group Formation Based on Learner Profile and Context. *International Journal on E-Learning, 5*(1), 19-24. Chesapeake, VA: Association for the Advancement of Computing in Education (AACE). Retrieved July 30, 2023 from https://www.learntechlib.org/primary/p/21767/.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*(4), 339–343. https://doi.org/10.1016/j.learninstruc.2009.08.003

Uchida, K. (2016). A Study on Pairwork Member Formation Methods in Information Literacy Education. *Aichi Prefectural University, Graduate School of Information Science and Technology, Doctoral Dissertation [published in Japanese].*

Kinjo, H. (2018). The Functioning of Imaginative Thinking Ability in Problem-Solving in Music Making in Elementary School Classes: Focusing on Heterogeneity in a Learning Group [published in Japanese]. *Journal of the Study of School Music Educational Practice*, *22*, 13-24. https://doi.org/10.19005/ssmep.22.0_13

Livingstone, S. R., & Russo, F. A. (2018, April 5). *The Ryerson Audio-Visual Database of emotional speech and Song (RAVDESS)*. Zenodo. https://zenodo.org/record/1188976

Sato, N., & Obuchi, Y. (2007). Emotion recognition using Mel-frequency cepstral coefficients. *Journal of Natural Language Processing, 14*(4), 83–96. https://doi.org/10.5715/jnlp.14.4_83