

# Investigating the Impression Effects of a Teacher-Type Robot Equipped a Perplexion Estimation Method on College Students

Kohei Okawa<sup>\*1</sup>, Felix Jimenez<sup>2</sup>, Shuichi Akizuki<sup>3</sup>, Tomohiro Yoshikawa<sup>4</sup>

<sup>1</sup>Graduate School of Information Science and Technology, Aichi Prefectural University, 1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, Japan

<sup>2</sup>School of Information Science and Technology, Aichi Prefectural University, 1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, Japan

<sup>3</sup>School of Engineering, Chukyo University, 101-2 Yagoto Honmachi, Showa-ku, Nagoya, Aichi, 466-8666, Japan

<sup>4</sup>Faculty of Medical Engineering, Suzuka University of Medical Science, 1001-1, Kishioka, Suzuka-city, Mie, 510-0293, Japan

## ARTICLE INFO

Article history:

Received: 30 March, 2023

Accepted: 08 July, 2023

Online: 25 July, 2023

Keywords:

Deep Learning

The Educational support robot

The Impression

Collaborative learning

## ABSTRACT

In recent years, the adoption of ICT education has increased in educational settings. Research and development of educational support robots have garnered considerable interest as a promising approach to inspire and engage students. Conventional robots provide learning support through button operations by the learners. However, the frequent need for button operation to request support may lead to a tedious impression on the learner and lower the efficiency of the learning process. Therefore, in this study, we developed a Perplexion Estimation Method that estimates the learner's state of perplexity by analyzing their facial expressions and provides autonomous learning support. We verified the impact of a teacher-type robot (referred to as the proposed robot) that autonomously provides learning support by estimating the learners' perplexity states in joint learning with university students. The results of a subject experiment showed that the impression of the proposed robot was not different from that of the conventional robot. However, the proposed robot demonstrated the ability to provide optimal support timing compared to the conventional robot. Based on these results, it is expected that the utilization of the perplexion estimation method with teacher-type robots can create a learning environment similar to human-to-human interaction.

## 1 Introduction

This paper is an extension of the one presented at SCIS [1]. In this conference, we presented the results of the life quality and the timing of support given to the learner by a supervised robot equipped with the proposed puzzling estimation method. In addition to these results, this paper provides additional analytical information on the results of subject experiments.

In recent years, the introduction of ICT education has become increasingly active in the field of education. The introduction of ICT education using educational big data that collects the learning status of individual students is being promoted in order to realize "fair, individualized and optimized learning that leaves no one behind" for children who have difficulty learning with other children due to reasons such as not attending school, children with developmental disabilities, and other children who are becoming increasingly

diverse [2]. By utilizing information technology, it is possible to provide individualized educational support. In this ICT education, collaborative learning with robots that have a "presence" in the real world is shown to be effective in creating a learning environment where people can learn from each other, as well as in stimulating interest in learning [3]. It has also been reported that the robot's advice is superior to that of an on-screen agent [4]. Therefore, we believe that educational support by robots is more effective than on-screen agents in educational settings. For these reasons, research and development of robots that can play an active role in educational settings (hereinafter referred to as "educational support robots") has been attracting attention in Japan and abroad [5].

Educational robots include "teacher-type robots" that instruct learners like a teacher. The role of a supervised robot in conventional research is to teach the learner how to solve a problem [6]. For example, Yoshizawa et al, proposed a supervised robot that

\*Corresponding Author: Kohei Okawa, Aichi Prefectural University, 1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, JAPAN, 0561-76-8826, im222002@cis.aichi-pu.ac.jp

switches learning support according to the number of correct answers given by the learner. Experimental results showed that the robot that switches the learning support has the potential to provide high learning effectiveness for university students.

However, conventional supervised robots have a problem in that they cannot autonomously provide learning support in response to the learner's state of being unable to solve a problem (hereafter referred to as "perplexion"). For example, Yoshizawa et al.'s teacher-type robot provides learning support by pressing a button of the learner (hereafter referred to as a "conventional robot"). The learner presses a button every time he/she needs learning support.

Conversely, enhancing the provision of support in learning systems is a critical focus in educational psychology research. In an environment where learners can request unlimited support, excessive use of hints [7] and search for hint patterns [8] occur regardless of the need for support. The use of artificial intelligence techniques has been reported to be effective in preventing these problems [9].

Therefore, the goal of this study creates a learning environment where a teacher-type robot provides autonomous learning support without the need for learners to press buttons, using deep learning techniques. We believe that it is effective for a teacher-type robot to estimate the learner's state of perplexion and autonomously provide learning support at the most appropriate timing, instead of providing learning support at the push of a button by the learner. Furthermore, when estimating the learner's state of perplexion, it is important to ensure that no burden is placed on the learner. We believe that this will prevent the learner from becoming dependent on the assistance and will enable smooth interaction with the robot, thereby improving the effectiveness of the teacher-type robot on the learner.

Regarding the learner's state estimation, Matsui has conducted previous research on autonomous learning support through perplexity state estimation [10]. Matsui's research attempted to estimate the learner's mental state by combining biometric devices, back-and-forth facial movements, and mouse movements. By using biological signals, we can obtain a state that is closer to the learner's raw data, which allows us to accurately estimate the perplexion. On the other hand, when using measurement equipment, the burden on the learner is large, and the data obtained is likely to contain noise in the real perplexity data. In the case of back-and-forth facial movements and mouse movements, the influence of the learner's posture and thinking habits is considered to be significant. Few studies have focused on estimating perplexion solely based on facial expression, despite the existing research on combining it with biometric signals.

In this study, we focus on research on human facial expression recognition [11] to provide autonomous learning support through teacher-type robots. In particular, methods based on deep learning have been widely used in research on facial expressions, and have shown high performance in image recognition and image classification Convolutional neural networks (hereafter, this is called CNN) have been proposed [12]. However, conventional research on facial expression recognition has focused only on the seven basic emotions of anger, disgust, fear, happy, sad, surprise, and neutral (hereafter referred to as the seven basic emotions), and has not focused on perplexion.

Therefore, in this study, we constructed the proposed method by transfer learning, using the seven basic emotion estimation methods of [13] as a base model. The proposed method is a perplexion

estimation method that classifies two classes of perplexion state and non-perplexion state. In this study, we first extended the seven emotion estimation method to the eight emotion estimation method, including the perplexion state (67% recall).

Then, we conducted a subjective experiment on the impression effect of a teacher-type robot equipped with the eight emotion estimation method (hereinafter referred to as "the eight emotion robot") and discovered that it elicited a similar level of the impression as a conventional robot [14]. On the other hand, the accuracy of the robot was not sufficient, as some subjects commented in a questionnaire that "the timing of support is too fast" and "the robot is noisy when it repeatedly speaks. To solve this problem, we constructed a proposed method to improve the estimation accuracy (88% recall). Nevertheless, we have been unable to make a direct comparison between the impressions and support timing of the robot equipped with the proposed method and the eight emotion robot.

In this paper, we verify the impression effect on university students and the support timing of joint learning with a teacher-type robot that provides learning support autonomously (hereinafter referred to as "the proposed robot") equipped with the proposed method specialized for estimating the perplexion on a teacher-type robot. In the experiment, we will conduct a comparison between the proposed robot and the eight emotion robot.

This paper begins with a description of the proposed method in Chapter 2. Then, in Chapter 3, we verify the impression that the proposed robot gives to the learner and the timing of support through experiments with participants. Chapter 4 discusses the results and Chapter 5 summarizes them.

## 2 Perplexion Estimation Method

Eight emotion estimation methods are designed to estimate the seven basic emotions together with the state of perplexion. However, including the estimation of the eight emotions reduces the accuracy of the estimation. Since the primary focus during learning is to determine whether the state is perplexed or not, estimating the eight emotions becomes unnecessary. In this chapter, we have developed a method for estimating perplexity using transfer learning, specifically focusing on estimating only the perplexed and non-perplexion states.

### 2.1 How to collect perplexed facial expression data

Perplexed state data were collected by capturing learners' facial expressions (at a resolution of 1920 and 1080 pixels, 30 frames per second) while they interacted with software related to the Technology Passport Examination (IT Passport) or software dealing with mathematical graphics difficulties and the Computer Aptitude Battery (CAB) [15] provided by SHL Japan. The participants in this study were university students, and the learning software was specifically designed to present them with moderately difficult problems that required careful thinking, based on their university lectures and high school mathematics knowledge.

The software presented a question with a hint button underneath. Participants were informed that they could press the button as many

times as they wanted to get pedagogical help. This facilitated the annotation of facial expression data during puzzles.

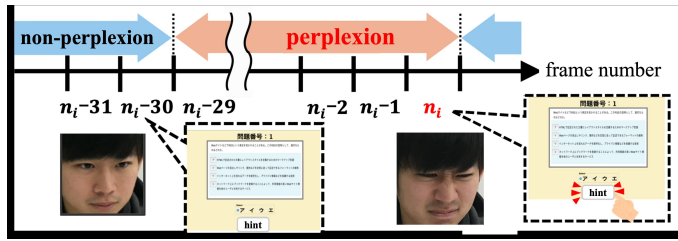


Figure 1: Definition of perplexion

In this study, facial expression data captured within 1 second (frames  $n_i - 29$  at 30 frames per second) before the  $i$ -th button press (frame  $n_i$ ) during learning with the learning software were categorized as perplexed state data, while other facial expression data were classified as non-perplexed state data (Fig. 1). The decision to define facial expression data up to 1 second ago as perplexed state data was based on the assumption that the expression of perplexity would be prominently evident during that period. Including facial expressions starting from 2 seconds ago may introduce variability in the strength of the puzzled state, leading to potential errors. Therefore, in this study, perplexed facial expressions are defined as the facial expressions that occurred 1 second before. The non-perplexed state is defined as the data collected from the frame immediately following the  $i$ -th button press (frame  $n_{i+1}$ ) to the frame just before the onset of the perplexed state at the time of the  $i$ -th button press (frames  $n_{i+i}$  to  $n_{i+i} + 30$ ).

## 2.2 Method Overview

In the field of deep learning, when the dataset available for training is small, transfer learning [16] is often utilized to leverage features learned by pre-existing models. In our study, in order to effectively capture subtle changes in facial expressions such as perplexion and non-perplexion, we employed seven basic emotion recognition models based on the FER2013 dataset as the base models (referred to as the "base model" in Fig. 2).

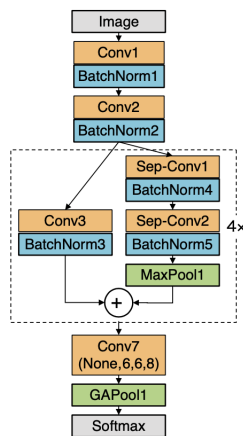


Figure 2: CNN model configuration

Specifically, we utilized the emotion recognition model using CNN developed by O. Arriaga et al [13]. By adapting the base model's features to focus on facial expressions, we anticipated an enhancement in the accuracy of perplexion estimation. Consequently, we constructed a dedicated model for perplexion estimation. In this model, the final layer generates the likelihood of two classes based on the perplexion state and the non-perplexion states. The perplexion estimation method of this paper was constructed using only facial expression data of college students. Therefore, the present method is specific to college students. To construct the perplexion estimation method, we utilized 36 face images out of the total of 52 collected in Section 2.1 as training data for perplexion/non-perplexion states. Of these, 26 images were used as training data (teacher data), while 10 images were kept aside for testing. Furthermore, since the method involves the binary classification of perplexion and non-perplexion states, we also included data from the FER2013 dataset representing other emotions into the non-perplexion data category (Table 1). Table 2 presents the number of data samples obtained from the aforementioned datasets.

The estimation of perplexion state is performed by inputting the face image to be estimated into a pre-trained model. The output of the model, denoted as  $y_c$ , represents the probability distribution over each class, with  $c$  indicating corresponding perplexion or non-perplexion state.

$$\arg \max_{1 \leq c \leq 2} y_c = 2 \tag{1}$$

Therefore, the state of perplexion is determined when the value of the formula (1) is "2", while the state of non-perplexion corresponds to all other cases. The process of estimating the perplexion state, referred to as "perplexion state estimation," commences after an average duration of time, denoted as  $x$ , has elapsed since the learner requests a hint. This approach prevents erroneous recognition that may occur when detecting, for example, a furrowed brow immediately upon the learner encountering a problem.

Table 1: Breakdown of the learning data.

	breakdown
Perplexion	Perplexion data collected
Non-Perplexion	Basic seven emotions from FER2013 Non-Perplexion data collected

Table 2: Configuration and number of learning data

Data Name	Data source	Number of data
anger	FER-2013	1997
disgust	FER-2013	218
fear	FER-2013	2048
happy	FER-2013	3607
sad	FER-2013	2415
surprise	FER-2013	1585
neutral	FER-2013	2482
non-perplexion	By Section 2.1	3504
perplexion	By Section 2.1	3475

To illustrate, the perplexion state estimation initiates after 30 seconds if the average time is set to be 30s.

### 3 participant experiment

#### 3.1 Robot overview

For the experiment, we utilized Tabot (Figure 3), a tablet-type robot with a tablet serving as its head, capable of displaying agents and expressing different facial expressions. Tabot's body consists of 3 degrees of freedom (DOFs) in the neck, 5 DOFs in an arm, and 1 DOF in the legs, for a total of 14 DOFs that allow different body movements.

In this experiment, as shown in Figure 3, a camera was installed on the robot's head to capture facial expressions during the learning process. The camera records the student's facial expressions as they look at the tablet screen during the learning process. The camera footage is regulated and processed by a processing PC. If a state of perplexion is detected from the captured facial expression data, an instruction is sent to Tabot via Local Area Network communication. The camera has a resolution of 640 and 480 pixels and a frame rate of 30 frames per second.

#### 3.2 Learning system

Participants study using the learning system shown in Figure 4. In the pre-training phase, subjects study, only with the learning system installed on their PCs, while in the collaborative phase, they study with the learning system displayed at the bottom of the tablet. The learning system displays the screen shown in Figure 4(a) when the subject proceeds to study. In Figure 4(a), there is a question and a button below the question that provides a hint. We informed the participants that they could press the button as many times as they wanted and that they would receive a hint by pressing the button. This makes it possible to label the facial expression data in the perplexion. However, the hint button is installed only in the pre-training phase, and is removed in the collaborative learning phase. When an answer is given on the screen shown in Figure 4(a), the user is taken to the screen shown in Figure 4(b). The system repeats this process for the number of questions to reach the final screen (Figure 4(c)).

The learning system comprises two types of challenges that are designed to involve a puzzle-like element or spark.



Figure 3: Tabot and camera

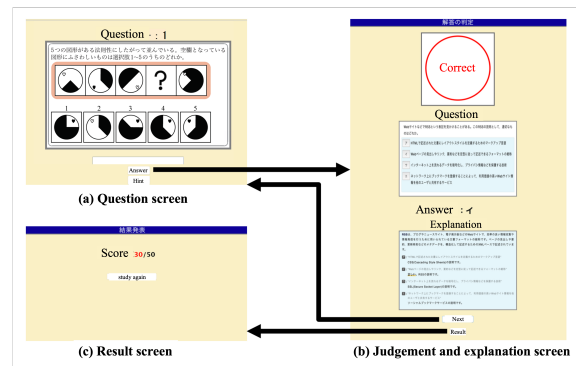


Figure 4: Learning system

These challenges consist of mathematical figure problems and Computer Aptitude Battery (CAB) problems provided by SHL Japan [15]. The mathematical figure problems can be effectively tackled by utilizing auxiliary lines in the drawing process, while the CAB problems involve identifying patterns and regularities. By presenting these challenges that demand both inspiration and puzzle-solving skills, we created an environment in which participants were consistently perplexed.

Table 3: Experimental information

Item	Contents
Property	Undergraduate and graduate students
The number of participants	34
Male:Female	20:14
Period	2020/9/14 11/30 2021/5/6 6/30 2022/2/17 3/17

#### 3.3 Experimental procedure

Experiments were conducted in which undergraduate and graduate students and robots learned together. Experimental information is summarized in the table 3 below.

In the experiment, we compare two groups of teacher-type robots equipped with different emotion recognition models. The groups to be compared are the eight emotion group equipped with the eight emotion estimation method and the proposed group equipped with the perplexion estimation method.

The information for the groups is shown in Table 4. For the proposed group, the estimated start time is set to 106 seconds after the first press of the hint button. This is because the average time elapsed until the first press of the hint button is 106 seconds during the experimental period up to 2021. This experiment is divided into three-time periods due to the number of participants. The proposal group was conducted in 2022, and the eight emotion group was conducted in the other dates.

The experimental procedure is shown below.

**Step 1 (Pre-learning)** Participants learn with a computer-based learning system. The learning time was about 60 minutes.

Table 4: Group information

	Eight emotion group	Proposed group
emotion recognition model	Eight emotion estimation method	Perplexion estimation method
hint provide condition	When estimating the perplexion	
Support frequency limitation	No limit	two times
hint button	nothing	
Estimated start time	Immediately after the problem is presented	After 106s

**Step 2** To reduce the learning effect, we allowed a one-week interval. In addition, the groups were assigned so that the number of times, the participants pressed the hint button in the pre-learning period would be as equal as possible.

**Step 3 (Collaborative learning)** Participants learned with a robot from the group to which they were assigned. The learning time was about 60 minutes.

**Step4(Survey)** Immediately after the collaborative learning the participants answer a survey.

### 3.4 Evaluation index

We employed the "Godspeed Questionnaire [17, 18]," a questionnaire methodology designed for the subjective evaluation of human-robot interaction. One of the questionnaire items utilized in our study was focused on assessing "Animacy." Additionally, we used the questionnaire to evaluate the timing of the robot's assistance. We selected these evaluation indices based on their perceived significance in the context of teacher-type robots. Previous research suggests that users tend to be more emotionally engaged and influenced by objects exhibiting animacy [19]. Therefore, a higher level of animacy in the robot may lead to a greater receptiveness to advice provided by the teacher-type robot. Furthermore, the autonomous timing of the robot's assistance is crucial for creating a more human-like learning environment. Each adjective pair in Godspeed Questionnaire was rated on a 5-point scale, and we quantified each pair on a scale of 1 to 5, with the positive adjective side receiving a higher score. We defined the average score of six animacy items as "animacy" and compared it across different groups. Figure 5 illustrates the questionnaire pertaining to support timing.

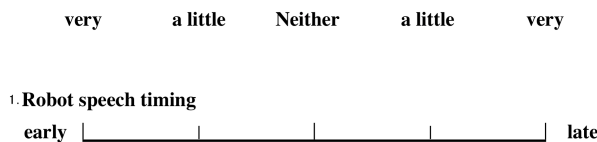


Figure 5: Questionnaire on support timing

Table 5: Scoring Procedure

survey	score
3	3
2 or 4	2
1 or 5	1

To analyze the evaluation results, we employed a Student's t-test for animacy, assuming equal population variance across the groups. Regarding the timing of support, two types of analyses were performed. The first analysis involved testing the population mean when the population variance was unknown, using "3" for "neither" as the optimal timing criterion. The second analysis utilized a Student's t-test with the timing questionnaire as the scoring criterion. Details of the scoring procedure can be found in Table 5. The significance level was adjusted to  $p < 0.0125$  using the Bonferroni correction to account for multiple comparisons, as four tests were conducted. The tests included comparisons of animacy, the average support timing between the eight emotion group and the proposed group, and comparison of support timing scores.

### 3.5 Result

Figure 6, 7, and 8 show the mean values of lifeliness and support timing, the mean values when scoring, and the test results. The proposed group had higher animacy than the eight emotion group. The test results showed no significant differences. These results suggest that the animacy that participants felt toward the robot was the same in each group of robots.

The support timing results for the "eight emotion group" and the "proposed group" were both found to be early and close to the criterion value of "3". The statistical analysis revealed a significant difference between the eight emotion group and an insignificant difference between the proposed group. This indicates that the eight emotion group was perceived as providing support too early, while the proposed group offered support at an appropriate timing. Moreover, employing the perplexion estimation method for support timing could lead to optimal timing of assistance. The results of scoring the timing of support were higher in the suggestion group than in the eight-emotion group.

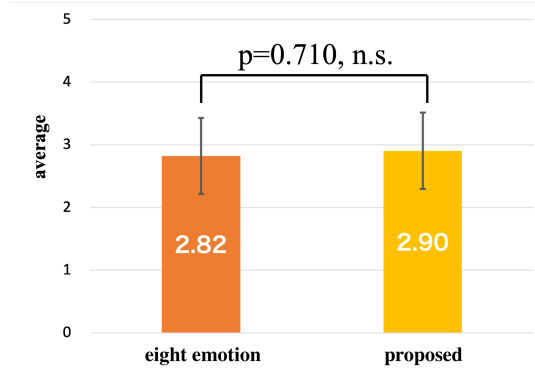


Figure 6: Result of Animacy

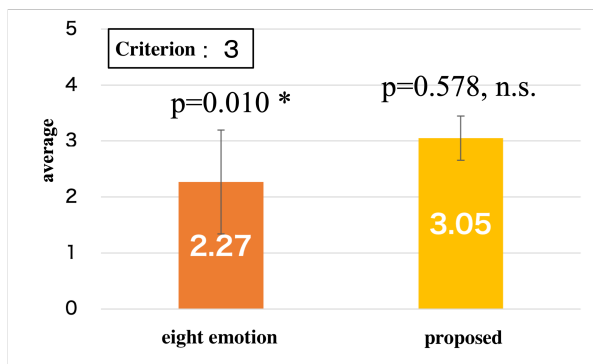


Figure 7: Result of Support timing

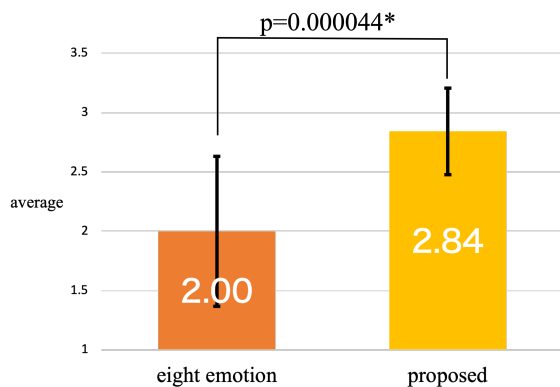


Figure 8: Result of support timing score

The results of the test showed that there was a significant difference. Therefore, the results indicated that the proposed group was better in support timing.

## 4 discussion

The experimental results indicated that there was no significant difference in animacy, suggesting that the learners had similar impressions regardless of the group they were in. However, a significant difference was found in the support timing between the eight emotion group and the proposed group. In the one-group test based on the criterion value of “3”, the eight emotion group showed a timing that was too early for the learners, while the proposed group provided support at the optimal timing. Furthermore, a significant difference was also found in the test when the questionnaires were converted into scores. These differences will be discussed in the following sections, along with the results of each questionnaire.

Animacy was higher in the proposed group than in the eight emotion group. We believe that this may be related to the number of speech utterances during collaborative learning. To investigate this, we examined the number of speech utterances for six participants in the eight emotion group and nineteen participants in the proposed group. The average number of utterances per participant was 29 for the eight emotion group and 0.2 for the proposed group. These results suggest that too many speech utterances may impair Animacy.

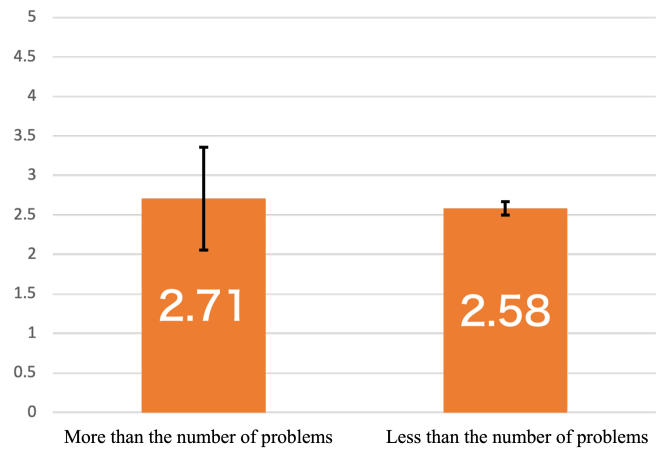


Figure 9: Animacy when divided by the number of support (The eight emotion group)

To further explore this, we investigated the Animacy scores of participants in the eight emotion group whose the number of speech utterances exceeded the number of questions. Four out of the six participants exceeded the number of questions, and we calculated the average Animacy scores of these four participants and the other two participants (Figure 9). Contrary to our expectations, Animacy was higher for those who received more support. However, due to the small number of participants, we cannot conclude with certainty that there is a clear trend. One possible explanation is that repeated utterances made the participants feel that there was a response.

In the Animacy section, there is a “responsiveness” column. The average score of the four respondents who received more support was 3.5, while that of the two respondents who received less support was 2. This result also suggests that a higher number of speech utterances may influence Animacy, as the respondents feel more responsive. However, the average response to the questionnaire regarding the timing of support was 1.75 when the number of times support was high and 2 when the number was low. Based on the open-ended responses from participants who had a high frequency of speech utterances, some of them expressed that there were “excessive number of hints” and they felt “confused by the continuous stream of advice”. Therefore, we believe that an excessive number of speech utterances does not have a positive effect.

Similarly, in the proposed group, the mean score of Animacy was calculated for four of the 19 participants for whom learning support was confirmed and for 15 participants for whom it was not confirmed (see Figure 10). Animacy scores were higher for the four participants for whom learning support was confirmed. Additionally, as in the eight emotion group, we calculated the mean score for each “responsive” item. The mean score for the four participants for whom learning support was confirmed was 4, and for the 15 participants for whom learning support was not confirmed, it was 3.3. Based on the analysis of the eight emotion group, we believe that having learning support is better than not having it, but that frequent learning support, such as continuous speech, may not give a good impression.

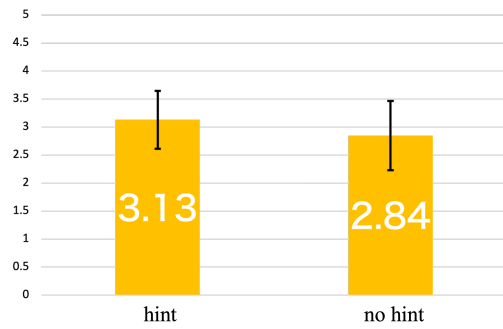


Figure 10: Animacy when divided by the number of support (The proposed group)

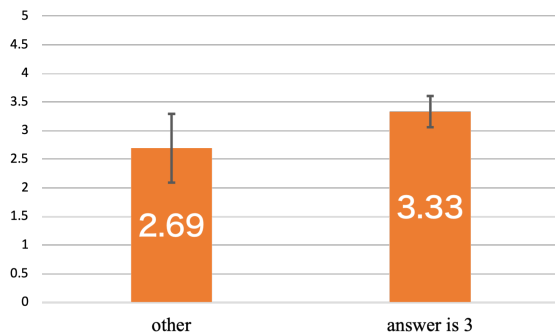


Figure 11: Animacy by support timing answers (eight emotion group)

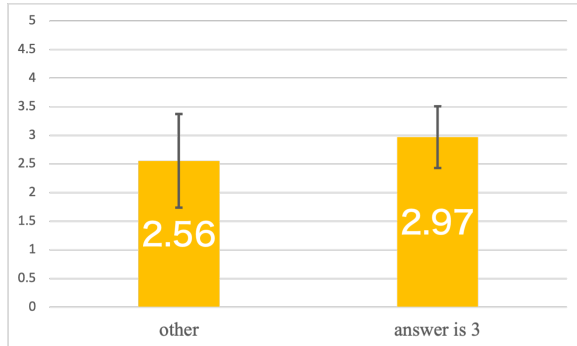


Figure 12: Animacy by support timing answers (proposed group)

We believe that the timing of support is an essential factor in autonomous support provided by a teacher-type robot. If support is given too early, learners may feel frustrated that it is offered at a time when they do not need it, whereas if it is provided too late, it may cause learners to feel frustration at not receiving support and not being able to solve problems. The experimental results showed a significant difference between the eight emotion group and the ideal timing of “3”. Therefore, it is thought that the participants felt that the support was premature. However, no significant difference was observed in the proposed group. This suggests that the support timing used in the proposed group may be the optimal timing. The test of scoring the timing of support showed a significant difference, with the proposed group having a higher score than the eight emotion group, indicating that the proposed group was superior. However, as seen in the previous data, the eight emotion group tends to talk

too much, resulting in a lower score. Similarly, the proposed group may not have been judged as early or late because they provided support significantly less frequently. Therefore, the results suggest that the perplexion estimation method used to solve this problem led to good results, as the evaluation of too many utterances was low.

Animacy and timing results suggest that teacher-type robots that provide appropriate timing and frequency of support can improve Animacy. Figures 11 and 12 show the average Animacy scores of participants who answered “3” in the questionnaire about support timing, and those who answered “other.” Three participants in the eight emotion group and 16 participants in the proposed group answered “3” for support timing, and in both cases, those who answered “3” had higher scores. These results indicate that refining support timing is effective in improving Animacy. To achieve this, we believe that the accuracy of the perplexity estimation method needs improvement. The current method has a reproducibility of 88%, but its accuracy is 67%, which is inferior to the eight emotion estimation methods. We believe that improving accuracy is necessary to provide an environment that can estimate the learner’s perplexity more accurately without missing it.

## 5 Conclusion

Our study developed a deep learning-based method to estimate the learner’s state of perplexion. We conducted experiments to compare the performance of a robot equipped with our proposed method to a robot equipped with the conventional eight emotion estimation methods. The results of the participant experiment revealed that there was no significant difference in terms of animacy between the two groups. This indicates that our proposed method provided a similar level of animacy to the learners compared to the conventional robot. However, when it comes to support timing, the robot equipped with our proposed method demonstrated the ability to provide support at the optimal timing. Furthermore, when the support timing was scored, a significant difference was observed, indicating that the support timing of the robot equipped with the proposed method was better. In other words, the proposed method can potentially enable the realization of a collaborative learning environment between a teacher-type robot and a learner without the need for buttons. Overall, we believe that this study has demonstrated the potential for realizing a collaborative learning environment between a robot and a learner using the proposed perplexion estimation method.

In the future, our plan is to develop a robot suitable for actual learning environments for junior high school and high school students, and to evaluate the impression the proposed robot leaves on learners as well as its effectiveness in supporting learning. To achieve this, we intend to conduct participant experiments that evaluate both the impression of the robot and its impact on learning. Moreover, to create a more effective learning environment for the learners, it is essential to enhance the accuracy of the perplexion estimation method. To achieve this, we are considering two approaches. The first approach is to fine-tune the method by training it on the data collected from the planned experiments, so that it can accommodate a wide range of age groups. This will enable us to effectively support learners with youthful facial features, such

as middle and high school students, who are expected to be actual users of the system. The second approach is to treat the learning data as time series data, including changes in facial expressions. Currently, we only focus on the moments of perplexion. However, by capturing the trends of perplexion onset, we anticipate that we can improve the estimation accuracy to a greater extent.

**Conflict of Interest** The authors declare no conflict of interest.

## References

- [1] K. Okawa, F. Jimenez, S. Akizuki, T. Yoshikawa, "Effects of a Teacher-type Robot Using a Perplexion Estimation Method on College Students," in 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), 2022, doi:10.1109/SCISISIS55246.2022.10001988.
- [2] Ministry of Education, Culture, Sports, Science and Technology, "Measure to Promote the Use of Advanced Technology to Support Learning in the New Era (Final Summary)," 2019, doi:https://www.mext.go.jp/a\_menu/other/1411332.htm.
- [3] T. Kanda, "How a communication robot can contribute to education?" *Journal of Japanese Society for Artificial Intelligence : AI*, **23**(2), 229–236, 2008, doi:10.11517/jjsai.23.2.229.
- [4] K. Shinozawa, F. Naya, J. Yamato, K. Kogure, "Differences in Effect of Robot and Screen Agent Recommendations on Human Decision-Making," *International Journal of Human-Computer Studies*, **62**(2), 267–279, 2005, doi:10.1016/j.ijhcs.2004.11.003.
- [5] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, "Social robots for education: A review," *Science Robotics*, **3**, 2018, doi:https://doi.org/10.1126/scirobotics.aat5954.
- [6] R. Yoshizawa, F. Jimenez, K. Murakami, "Proposal of a Behavioral Model for Robots Supporting Learning According to Learners' Learning Performance," *Journal of Robotics and Mechatronics*, **32**(4), 769–779, 2020, doi:https://doi.org/10.20965/jrm.2020.p0769.
- [7] V. Aleven, K. R. Koedinger, "Limitations of student control: Do students know when they need help?" in *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, 2000, doi:10.1007/3-540-45108-0\_33.
- [8] J. A. Walonoski, N. T. Heffernan, "Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems," in *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 2006, doi:10.1007/11774303\_38.
- [9] I. Roll, V. Aleven, B. M. McLearn, K. R. Koedinger, "Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system," *Learning and Instruction*, **21**(2), 267–280, 2011, doi:10.1016/j.learninstruc.2010.07.004.
- [10] T. Matsui, "Estimation of Learners' Physiological Information and Learners' Mental States by Machine Learning and Its Application for Learning Support," *Transactions of Japanese Society for Information and Systems in Education*, **36**(2), 76–83, 2019, doi:https://doi.org/10.14926/jjsise.36.76.
- [11] M. Wimmer, B. A. MacDonald, D. Jayamuni, A. Yadav, "Facial Expression Recognition for Human- Robot Interaction - A Prototype," in *In Proceedings of the 2nd International Workshop on Robot Vision*, 2008, doi:10.1007/978-3-540-78157-8\_11.
- [12] T. Nishime, S. Endo, K. Yamada, N. Toma, Y. Akamine, "Feature Acquisition From Facial Expression Image Using Convolutional Neural Networks," *Journal of Robotics, Networking and Artificial Life*, **32**(5), 1–8, 2017, doi:http://dx.doi.org/10.2991/jrnal.2016.3.1.3.
- [13] O. Arriaga, P. G. Ploger, M. Valdenegro, "Real-time Convolutional Neural Networks for Emotion and Gender Classification," 2017, doi:10.48550/arXiv.1710.07557.
- [14] K. Okawa, F. Jimenez, K. Murakami, S. Akizuki, T. Yoshikawa, "A Survey on Impressions of Collaborative Learning with an Educational Support Robot Estimating Learners' Perplexion," in *Proceedings of the 37th Fuzzy Systems Symposium*, 171–176, 2021, doi:10.14864/fss.37.0.171.
- [15] SHL Japan products, 2020/12/06 accessed, doi:http://www2.shl.ne.jp/product/index.asp?view=recruit.
- [16] A. Gulli, S. Pal, Chokkan Deep learning: Python×Keras de aidea o katachi ni suru reshipi, O'Reilly Japan, 2018, doi:https://www.ohmsha.co.jp/book/9784873118260/.
- [17] C. Bartneck, D. Kulic, E. Croft, S. Zoghbi, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived, Intelligence, and Perceived Safety of Robots," *International Journal of Social Robotics*, **1**(1), 71–81, 2009, doi:https://psycnet.apa.org/doi/10.1007/s12369-008-0001-3.
- [18] T. Nomura, "Humans' Subjective Evaluation in Human-Agent Interaction(HAI)," *Journal of the Japanese Society for Artificial Intelligence*, **31**(2), 224–229, 2016, doi:https://doi.org/10.11517/jjsai.31.2.224.
- [19] B. J. Fogg, *Persuasive Technology: Using Computers to Change what We Think and Do*, Morgan Kaufmann, 2002, doi:10.1016/B978-1-55860-643-2.X5000-8.