Investigating Multimodal Predictors of Peer Satisfaction for Collaborative Coding in Middle School

Yingbo Ma, Gloria Ashiya Katuka, Mehmet Celepkolu, Kristy Elizabeth Boyer
University of Florida
{yingbo.ma, gkatuka, mckolu, keboyer}@ufl.edu

ABSTRACT

Collaborative learning is a complex process during which two or more learners exchange opinions, construct shared knowledge, and solve problems together. While engaging in this interactive process, learners' satisfaction toward their partners plays a crucial role in defining the success of the collaboration. If intelligent systems could predict peer satisfaction early during collaboration, they could intervene with adaptive support. However, while extensive studies have associated peer satisfaction with factors such as social presence, communication, and trustworthiness, there is no research on automatically predicting learners' satisfaction toward their partners. To fill this gap, this paper investigates the automatic prediction of peer satisfaction by analyzing 44 middle school learners' interactions during collaborative coding tasks. We extracted three types of features from dialogues: 1) linguistic features indicating semantics; 2) acoustic-prosodic features including energy and pitch; and 3) visual features including eye gaze, head pose, facial behaviors, and body pose. We then trained several regression models to predict the peer satisfaction scores that learners received from their partners. The results revealed that head position and body location were significant indicators of peer satisfaction: lower head and body distances between partners were associated with more positive peer satisfaction. This work is the first to investigate the multimodal prediction of peer satisfaction during collaborative problem solving, and represents a step toward the development of real-time intelligent systems that support collaborative learning.

Keywords

Collaborative Learning, Peer Satisfaction, Pair Programming, Multimodal Learning Analytics

1. INTRODUCTION

Collaborative learning benefits learners in numerous ways, such as enhancing critical thinking [31], developing social skills [29], and improving learning gains [32]. During collaborative learning, partners may bring different ideas to solve a problem, defend and evaluate their perspectives, and have a dynamic interaction with each other to produce a shared solution [18]. This relationship between partners can be a decisive factor for the success of the collaboration and positive team experience [9], and the partners' satisfaction toward each other can have a significant impact on their task performance [52] and learning outcomes [20]. Previous literature suggests that students' interactions may not be productive and they may face challenges with their partner [25, 6], which could discourage them from working with partners in the future [44]. In a classroom setting, teachers may not have the resources to detect whether the partners in a team have positive attitudes toward each other and enjoy working together. Therefore, it becomes even more important to develop intelligent and adaptive technologies to predict peer satisfaction during collaborative activities.

Despite the increase in the development of techniques and models to analyze students' interactions during collaborative learning [49, 45], there is no research on automatically predicting peer satisfaction during collaboration. Current studies that analyzed learners' satisfaction during collaboration have revealed important factors such as social presence (sense of being with each other [46, 27]), frequency and quality of team communication [28], and mutual trust between group members [53]). However, most of these post-hoc studies relied on manual approaches (e.g., analyzing post-study attitude survey [22] or open-ended questions [28]). On the other hand, multimodal learning analytics research has created new opportunities to automatically analyze learners' interactions from multiple modalities (e.g., speech, facial expressions, body gestures), and provide insights into the learning process from different dimensions [2]. For example, recent studies successfully classified critical facets of collaborative problem solving process with multimodal features (linguistic, acoustic-prosodic, facial expressions, and task context) derived from groups of learners' collaborative dialogues [48]. However, multimodal learning analytics has not yet been used to automatically predict peer satisfaction from learners' interactions.

Aligned with this motivation, our goal in this paper is to investigate the automatic prediction of peer satisfaction during collaborative learning. We specifically address the following two research questions (RQs):

- RQ 1: What are the most predictive unimodal features of peer satisfaction during collaboration?
- RQ 2: Does multimodal feature fusion improve peer satisfaction prediction compared to the best-performing unimodal model?

To answer these research questions, we analyzed audio and video data collected from 44 middle school learners who worked in pairs on a series of collaborative coding activities. After participating in coding activities, each learner reported their overall satisfaction with their partners. To answer RQ 1, we examined the performance of the following features extracted from learners' collaborative dialogues, including: 1) linguistic features indicating semantics from Word2Vec [35] and pre-trained BERT [15]; 2) acoustic-prosodic features such as energy and pitch extracted with openSMILE [17]; 3) eye gaze, head pose, and facial AUs extracted with OpenFace [1]; and 4) body pose extracted with OpenPose [3]. We followed a state-of-the-art methodology [50] that preserves the sequential nature of the features across the collaborative session.

The experimental results revealed two significant predictors. The first significant predictor was head position (x-axis), generated from OpenFace, which was the horizontal distance of a learner's head from the camera (located in the middle of two learners to collect video recordings). The second significant predictor was body key points (x-axis), generated from OpenPose, which was the the horizontal pixel location of a learner's eight upper body key points (e.g., nose, neck, and shoulders). These results indicated that learners who had lower head and body distances from their partners were more likely to receive higher peer satisfaction scores. To answer RQ 2, we evaluated the model performance of several early-fused multimodal features, and the results showed that the multimodal features investigated in this study did not significantly improve the prediction accuracy of peer satisfaction compared to the best-performing unimodal feature.

This study provides two main contributions: 1) we present the results from extensive experiments evaluating both a variety of predictive features and a selection of sequential models; 2) and we identify two interpretable and meaningful learner behaviors that can be predictive of peer satisfaction. To the best of our knowledge, this is the first study to investigate the automatic prediction of peer satisfaction with multimodal features extracted from learners' interactions.

The rest of the paper is organized as follows: Section 2 presents the related work; Section 3 describes the dataset used for this study; Section 4 details the features we investigated;. Section 5 elaborates on the peer satisfaction prediction models; Section 6 presents the experimental settings and results; Section 7 discusses the implications of experimental results; and finally, section 8 concludes the paper and discusses future work.

2. RELATED WORK

Interpersonal interactions and soft skills play an important role in students' learning experiences and teams' success during collaboration [11]. Previous research has emphasized that partners may have trouble while collaborating on a task together for a variety of reasons, and many social factors can have an impact on peer satisfaction. For example, So et al. [46] recruited 48 graduate students who collaborated on a healthcare project. They found that learners' perceived social presence and emotional bonding were important factors for peer satisfaction. Zeitun et al. [52] examined the relationship between team satisfaction and course project performance among 65 groups of students. They found that team satisfaction (toward partners and their collaborative work) were positively related to group performance only for American students, and there was no significant difference in the satisfaction and performance regarding gender. Katuka et al. [26] analyzed the relationship between dialogue act and peer satisfaction from 18 pairs of middle school students. They identified six sequences of dialogue acts (e.g., questions, clarifications) that were positively related to satisfaction. Despite the insights of peer satisfaction provided by the aforementioned studies, most of these studies relied on manual approaches (e.g., post-study attitude survey or open-ended questions), which does not enable the automatic prediction of peer satisfaction.

In recent years, there has been an increasing interest in using multimodal learning analytics (MMLA) techniques that combine multiple data streams (e.g., speech and spoken words [41], text message and facial expressions [14]) to analyze student collaborative interactions. For example, Spikol et al. [47] used MMLA to estimate the success of collaboration with face tracking, hand tracking, and audio recording. They found that distances between learners' hands and faces were two strong indicators of group performance, and lower distances indicated that it was more likely that successful collaboration occurred among students. Echeverria et al. [16] applied MMLA in a healthcare setting in which nurses collaborated in groups, with their audio, movement, and physiological data collected and analyzed. The authors demonstrated that integrating more sources of data multimodal data provided more contextual details of group activities during collaboration process. In another study, Liu et al. [30] used MMLA to understand learners' knowledge model refinement process during collaboration. They were able to better predict learners' knowledge models when they combined multiple data streams (i.e., audio, screen video, webcam video, and log files), which convey important contextual information about student learning. However, to the best of our knowledge, there is no research on automatic prediction of peer satisfaction using multimodal features during collaborative learning. Our study extends this body of MMLA research on learners' interactions. We investigate different modalities (linguistic, acoustic-prosodic, and visual) for automatically predicting peer satisfaction.

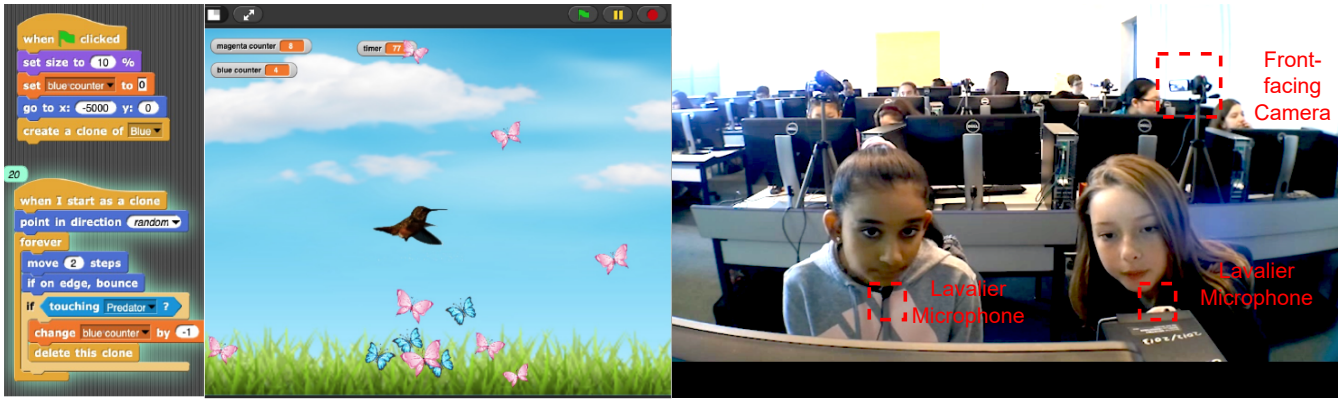

Figure 1: *Left*: A sample script created with Snap!. *Right*: Two middle school learners collaborating on a pair programming task. In the captured moment, the learner in the left side of the frame is the *driver* and the learner on the right is the *navigator*; their collaborative interaction is video-recorded with a front-facing camera and audio-recorded with each learner wearing a lavalier microphone.

3. DATASET

3.1 Participants and Collaborative Activities

Our dataset was collected from 44 learners in 7th grade classrooms in a middle school in the southeastern United States during two semesters (Spring and Fall 2019). Out of 44 learners, 29 (65.9%) identified themselves as females and 15 (34.1%) as males. The distribution of race/ethnicity of these learners included 41.3% self-reporting as White, 26.1% Asian/Pacific Islander, 19.5% Multiracial, 8.7% Hispanic/Latino, 4.3% Black/African American, and 1.9% Other. The mean age was 12.1 with ages ranging from 11 to 13.

The learners collaborated on a series of coding activities in which they practiced fundamental CS concepts such as variables, conditionals, and loops using Snap! block-based programming environment [7]. The learners followed the pair programming paradigm, in which each pair shared one computer and switched roles between the *driver* and the *navigator* during the science-simulation coding activity (Figure 1). The *driver* is responsible for writing the code and implementing the solution, while the *navigator* provides support by catching mistakes and providing feedback on the in-progress solution [4].

3.2 Data Collection and Text Transcription

The collaborative coding session of each pair was recorded at 30 fps in 720p through a front-facing detached camera, and each child wore a lavalier microphone without active noise cancelling. The audio was recorded by digital sound recorders with a sample rate of 48KHz. After the audio/video data collection process was finished, an online manual transcription service [42] generated the textual transcript for each pair. The transcripts included three pieces of information for each spoken utterance: (1) *Starting Time*, in the form of *hour*:*min*:*sec*; (2) *Speaker*, in the form of *S1* (the learner sitting on the left of the video) or *S2* (the learner sitting on the right); and (3) *Transcribed Text*. Each collaborative coding session took around 30 minutes. In total, the corpus included 12 hours and 18 minutes of audio and video recordings, with 10,265 transcribed utterances.

3.3 Peer Satisfaction Post Survey

After participating in the collaborative coding sessions, each learner completed a peer satisfaction post survey. To the best of our knowledge, there is no existing validated survey for peer satisfaction in the pair programming context, so we developed a 6-item survey based on previous surveys on peer satisfaction. Sample questions in the peer satisfaction survey included: "My partner answered my questions well", "My partner listened to my suggestions", and "My partner often cut my speech". Each of the six items in the survey was measured on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

Figure 2 (*Left*) shows the distribution of the peer satisfaction post survey responses from 44 learners. The distribution of the satisfaction scores shows that most of the learners agreed or strongly agreed that they were satisfied with the overall interaction with their partner. To determine whether to treat the six post-survey items as a single item or multiple items, we conducted a principal component analysis (PCA). The results of PCA suggested proceeding with only one derived outcome variable, which we refer to as *Satisfaction* score (the average score of six items). This derived outcome explains 52% of the variation across all six survey items, with an eigenvalue of 3.15. Figure 2 (*Right*) shows the distribution of the averaged *Satisfaction* score. The mean value of the *Satisfaction* score is 4.3 (SD=0.6) out of 5, with a maximum value of 5.0, and a minimum value of 2.2.

4. FEATURES

In this section, we introduce the feature extraction process from the audio (section 4.1), video (section 4.2), and language (section 4.3) modalities. Then we describe the feature padding process (section 4.4) that prepared the extracted features for model training. Table 1 shows the features this study investigated, and their corresponding dimensional details after the feature padding process.

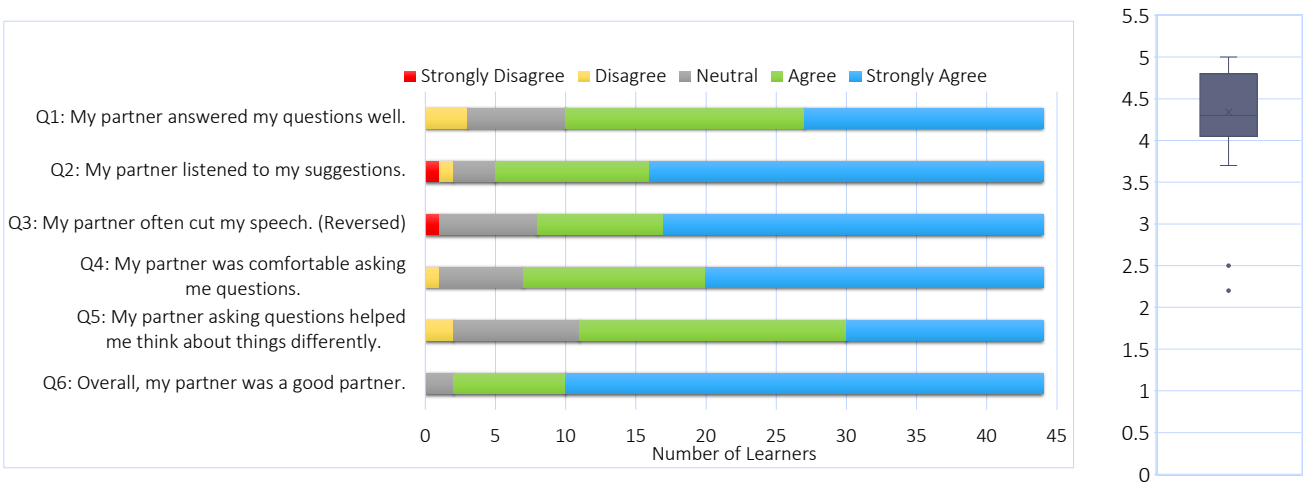

Figure 2: *Left*: distribution of peer satisfaction post-survey items from from 44 learners. *Right*: distribution of the continuous averaged *Satisfaction* score (mean = 4.3, SD = 0.6).

Table 1: Utterance-Level Features

Modality	Feature Name	Vector Dimension*
Audio	Loudness	638, 704, 990
	Pitch	580, 640, 900
	Shimmer	116, 128, 180
	Jitter	116, 128, 180
	MFCCs	928, 1024, 1440
Language	Word Count	1
	Speech Rate	1
	Word2Vec	4,200
	Pre-trained BERT	768
Video	Eye Gaze Directions	784
	Head Directions	294
	Head Position (x-axis)	98
	Head Position (y-axis)	98
	Head Position (z-axis)	98
	Facial AUs	3,430
	Body Key Points (x-axis)	784
	Body Key Points (y-axis)	784

* For every audio-based feature, the three vector dimensions resulted from different speech lengths (29s, 32s, and 45s) after applying three different silence removal thresholds (-6, -16, and -30 dBFS) respectively. For language-derived features, the maximum number of spoken words was 42. For video-derived features, the maximum time length of video segments was 49s.

4.1 Audio-based Features

Simple acoustic-prosodic features (e.g., sound level, synchrony in the rise and fall of the pitch) derived from audio have proven to be effective in predicting learners' engagement level [49] and estimating group performance on solving open-ended tasks. [47]. In our study, we extracted audio-derived features on the corresponding audio segment for each utterance. Because we only obtained the *Starting Time* of each utterance from the online transcription service, and not the *Ending Time*, the raw audio segments in our corpus also contain silence (background noise when a learner stops talking) that elapsed before the next utterance started. To mitigate the potential negative influence of this silence in our audio segments, we used *pydub.detect_silence*, a function in the *pydub* [38] library to detect the time of end-of-utterance in a given audio segment. The function required a pre-defined parameter: silence removal threshold (any audio lengths quieter than this will be considered as silence). For each raw audio segment, we used three different silence thresholds to produce three different audio segments: -6 dBFS (half of the audio's maximum level), -16 dBFS (default setting of the function), and -30 dBFS (low enough to avoid losing actual speech lengths).

After removing the potential silence contained in raw audio segments, we used openSMILE v2.2, an open-source acoustic feature extraction toolkit, for automatic extraction of the following five types of audio-based features within a 20-ms frame and a 10-ms window shift. The five categories of audio-based features are as follows:

1. **Loudness** measures the energy level of the signal. For each audio frame, 11 loudness-related features were extracted.

2. **Pitch** measures the frequency scale of a signal. For each audio frame, 10 pitch-related features were extracted.

3. **Shimmer** measures how quickly the loudness of the signal is changing. For each audio frame, 2 shimmer-related features were extracted.

4. **Jitter** measures how quickly the frequency of the signal is changing. For each audio frame, 2 jitter-related features were extracted.

5. **MFCCs** (Mel-Frequency Cepstral Coefficients) measures the shape of the signal's short-term spectrum. For each audio frame, 16 MFCCs-related features were extracted.

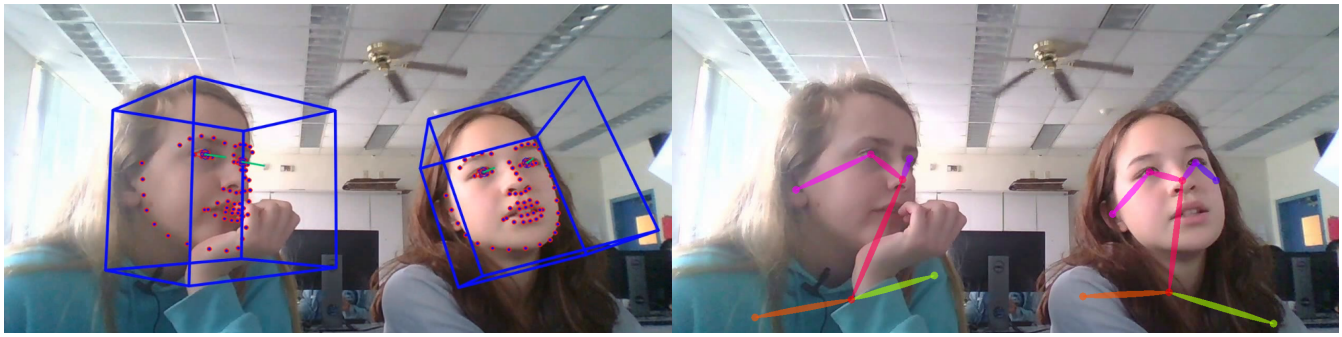

Figure 3: An example of the video-derived feature extraction process for both learners in a specific video frame. *Left*: eye gaze direction (green vectors), head pose (blue 3D bounding boxes), and facial AUs (recognized later) extracted with OpenFace. *Right*: upper body key points (e.g., nose, neck, and shoulders) extracted with OpenPose.

4.2 Language-based Features

Linguistic features extracted from spoken utterances have been used to model collaborative problem solving skills and predict collaborative task performance [37]. In our study, multiple commonly used statistical and semantic linguistic features were extracted for each spoken utterance. The four categories of language-derived features are as follows:

1. **Word Count** For each utterance, word count was calculated as the number of words.

2. **Speech Rate** For each utterance, speech rate was calculated as the number of words divided by the number of elapsed seconds in the utterance, to produce words per second.

3. **Word2Vec** is a semantic method which learns word associations from the text, and groups similar words together in a vector space based on their semantics. We train our Word2Vec model with *gensim*, an open-source natural language processing library. The default settings of parameters were used, in which the dimension of each word embedding was set to 100, with a sliding window size of 5.

4. **Pre-trained BERT** is a language model trained on a large amount of data (e.g., texts from Wikipedia and books) in a self-supervised way. Similar to Word2Vec, BERT represents semantics of words in a vector space. In this study, we used the BERT-base-uncased model, which is a publicly available BERT model trained only on English texts with the hidden size of 768. With this pre-trained BERT, we generated one 768-dimensional vector for each utterance.

4.3 Video-based Features

A variety of features generated from video modality have been investigated in prior literature modeling collaborative problem solving. For example, eye gaze has proven effective in evaluating learners' attentiveness [43, 24] and learning performance [5, 40]; head pose has also been used for assessing learners' collaborative problem solving competence [13]; facial action units (AUs) have been used to measure both individual learners' tutoring outcomes [21] and interaction level during collaborative learning [33]. Body pose has been used for analyzing learners' engagement level [39] and modeling collaborative problem solving competence [13]. In our study, video-derived features were extracted from the corresponding raw video segment of each utterance. We used the OpenFace v2.0 facial behavior analysis toolkit and OpenPose v1.7 body key points detection toolkit to extract the following four categories of video-based features (See Figure 3):

1. **Eye Gaze Direction** refers to the direction in which an eye looks. For each detected face per video frame, 8 eye gaze direction-related features were extracted. They included 3 eye gaze direction vectors (x direction, y direction, and z direction) for each eye, and 2 eye gaze directions in radians averaged for both eyes.

2. **Head Pose** refers to head position and direction. For each detected face per video frame, 6 head-related features were extracted with OpenFace, including three head position vectors (x direction, y direction, and z direction) representing the location of the head with respect to the camera in millimeters, and three head direction vectors in radius with respect to the camera. Since the front-facing camera was located in the middle of two learners during the data collection process, positive values of the x direction vector and the z direction vector indicate that the learner is sitting on the right side of the video and away from the camera, and vice versa. The head position features used in our study were the absolute values of the x, y, and z direction vectors, representing the spatial location of each learner's head from the camera.

3. **Facial AUs** refer to the movements of an individual's facial muscles. For each detected face per video frame, 35 facial AU-related features were extracted with OpenFace, including 17 facial AU intensity features (ranging from 0 to 5), and 18 facial AU presence features (0-absence or 1-presence).

4. **Body Pose** refers to the location of each joint (e.g., neck, shoulders) of the human body, which are known as key points that can describe a person's pose. For each learner appearing in each video frame, the 2D locations (x direction and y direction) of 8 body key

points, measured in pixels, were extracted with OpenPose. These included the locations of each learner's eyes, nose, neck, and shoulders. OpenPose supports real-time detection of 25 full body key points (hand, facial, and foot key points); however, since our video recordings only captured learners' upper bodies, OpenPose was not able to detect the locations of some body points such as hand and foot. Therefore, only 8 body key points related to learners' upper bodies were extracted and used in this study. Because the resolution of our cameras was 720p, the maximum pixel value of body key points generated from OpenPose was 1280 pixels in the x direction, and 720 pixels in the y direction.

4.4 Feature Padding

Spoken utterances naturally vary in time length, and feature padding is an important step for ensuring the uniform size of model inputs before training machine learning models. We averaged the audio-based and video-based features across a small non-overlapping time window because they were extracted on the frame level. Following the feature aggregation methods used in prior works [47, 49], in which the average time windows of 500 ms and 1000 ms were chosen respectively, we selected the time window of 500 ms. We did not choose a longer window because audio-based features (e.g., pitch) could vary over a longer duration, which would lead to losing fine-grained details. Finally, post padding (adding zeros to the end of vectors) was applied on each averaged feature vector with the maximum time length (29s, 32s, and 45s) for different silence removal thresholds. For the Word2Vec-based feature, word embeddings were concatenated to form one feature vector for each utterance. Then, post padding was applied to the Word2Vec-based feature vector and the BERT-based feature vector with the maximum number (42) of spoken words.

5. PREDICTION MODELS

Figure 4 depicts the architecture of our peer satisfaction prediction model. For a collaborative coding session of a given pair (Learner A and Learner B), the model input is a session-level feature sequence $X = [x_0, x_1, ..., x_{N-1}]$ for Learner B, in which N is the number of spoken utterances from the learner. For each element in X, we used the early fusion method to generate utterance-level multimodal feature $x_t = [a_t, v_t, x_t]$ by concatenating unimodal audio-derived feature a_t, video-derived feature v_t, and language-derived feature l_t. Before concatenating unimodal feature vectors into a single multimodal feature vector, we applied z-score normalization to all the features by subtracting their mean value and dividing by their standard deviation.

Our prediction model contains two stages: feature learning and regression. In the feature learning stage, we followed the current state-of-the-art methodology [50] that preserves the sequential nature of dialogue to learn the input feature sequence X. The sequential model is a two-layer LSTM network with 128 units. We obtained a final 128-dimensional hidden state h_T from the sequential model. During the regression stage of the model, we used h_T as input, and fully connected layers to output a continuous estimated satisfaction score $\widehat{y}$, in order to approximate the actual *Satisfaction* score y rated by Learner A.

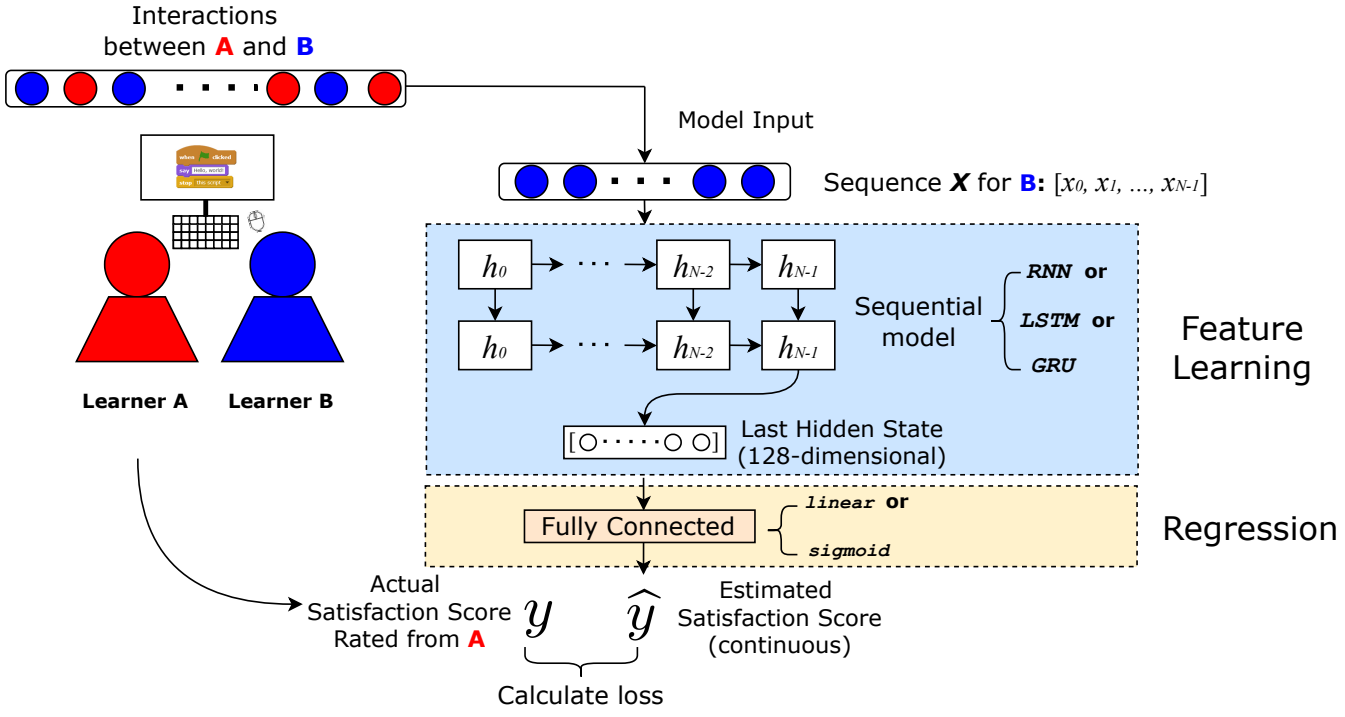

Figure 4: Architecture of the prediction model. For unimodal modeling, x_t $(0 \leqslant t \leqslant N-1)$ is a unimodal feature vector (audio a_t, video v_t, or langauge l_t). For multimodal modeling, x_t is a subset of an early-fused vector $[a_t, v_t, x_t]$ (normalized).

Table 2: Regression results of unimodal models. Two highlighted features: Head Position (x-axis) and Body Key Points (x-axis), significantly reduced the MAE compared to the baseline feature ($p\text{-}value < .05$).

Modality	Unimodal Feature	***MAE***	***p-value*** ($\widehat{y}_{base}$ ***and*** $\widehat{y}$)	R^2 (y *and* $\widehat{y}$)
	Baseline	0.1953	—	0.07
Audio	Loudness	0.1981, 0.1790, 0.1796	0.31, 0.19, 0.12	0.02, 0.05, 0.07
	Pitch	0.2073, 0.1902, 0.1881	0.25, 0.14, 0.15	0.01, 0.15, 0.03
	Shimmer	0.1895, 0.1794, 0.1713	0.12, 0.19, 0.42	0.04, 0.12, 0.20
	Jitter	0.1983, 0.1896, 0.1853	0.14, 0.31, 0.31	0.01, 0.08, 0.06
	MFCCs	0.2341, 0.2405, 0.2318	0.19, 0.19, 0.24	0.01, 0.03, 0.03
Language	Word Count	0.1794	0.43	0.07
	Speech Rate	0.1790	0.19	0.29
	Word2Vec	0.1751	0.09	0.06
	Pre-trained BERT	0.1789	0.06	0.08
Video	Eye Gaze Directions	0.1689	0.10	0.21
	Head Directions	0.1583	0.09	0.23
	Head Position (x-axis)	0.1402	0.03	0.68
	Head Position (y-axis)	0.1902	0.21	0.15
	Head Position (z-axis)	0.1640	0.09	0.25
	Facial AUs	0.1927	0.19	0.11
	Body Key Points (x-axis)	0.1376	0.03	0.64
	Body Key Points (y-axis)	0.1761	0.39	0.10

MAE: aggregated testing absolute error for all data samples. y: actual satisfaction scores. $\widehat{y}$: predicted satisfaction scores with each unimodal feature. $\widehat{y}_{base}$: predicted satisfaction scores with the baseline feature. R^2: another widely used metric to evaluate a regression task's level of goodness-of-fit.

Table 3: Regression results of multimodal models. None of the multimodal features significantly outperformed the baseline feature.

Multimodal Feature	***MAE***	***p-value*** ($\widehat{y}_{base}$ **and** $\widehat{y}$)	R^2 (y *and* $\widehat{y}$)
Baseline: Body Key Points (x-axis)	0.1376	—	0.64
Head Position (x-axis, z-axis)	0.1484	0.39	0.65
Head Position (x-axis), Head Directions	0.1355	0.17	0.68
Head Position (x-axis), Body Key Points	0.1367	0.10	0.68
Head Position (x-axis), Pre-trained BERT	0.1409	0.13	0.65

Recent research [50] has shown that the type of sequential model can play an important role for feature learning. Therefore, we also evaluated the performance of recurrent neural network (RNN) and gated recurrent unit (GRU) models to understand the influence of different sequential model architectures during feature learning. In addition, we evaluated the performance of two different output units, *sigmoid* and *linear* functions, to compare between linear and non-linear regression.

6. EXPERIMENTS AND RESULTS

6.1 Experimental Setups

We implemented the Python code[1] for our prediction models in Keras with a Tensorflow backend. We conducted five-fold cross-validation to train and validate the models. All labels (y) were normalized (ranging from 0 to 1) before the model training process because the *sigmoid* activation function was used to produce the predicted satisfaction scores $\widehat{y}$. We used Adam optimizer with the learning rate of $1 \times e^{-3}$ to train the prediction model, which was trained for up to 100 epochs. The mean absolute error (MAE) was calculated for the loss function. After five rounds of cross-validation, we aggregated the MAE of each fold during the model testing process.

[1] `https://github.com/yingbo-ma/Predicting-Peer-Satisfaction-EDM2022`

6.2 Investigating Unimodal Features

To identify predictive unimodal features, we compared the prediction accuracy of each unimodal feature with a randomly generated baseline feature. Followed a common method of generating uniform random baselines [12, 19], we used the Python function *random.uniform*(0, 1), which can be interpreted as white noise without any meaningful content. We then trained the model with the white noise to generate the random baseline results (error MAE_{base} and predicted scores $\widehat{y}_{base}$). This low baseline allows us to measure the extent to which each feature predicts the outcome better than random chance. Next, we trained the model with each of the unimodal features from Table 1, and generated corresponding MAE and $\widehat{y}$. A paired-samples t-test [34] between $\widehat{y}_{base}$ and $\widehat{y}$ checked whether adding that unimodal feature significantly reduced error compared to the random baseline. Table 2 shows the regression results of peer satisfaction prediction models trained on unimodal features.

For audio-derived features, the three values in each column (from left to right) resulted from different silence removal thresholds (-6, -16, and -30 dBFS). Although time-domain features (e.g., Loudness, Shimmer) performed better than frequency-domain features (Pitch, Jitter), as indicated by lower *MAEs*, the associated *p-values* showed that none of the acoustic and prosodic features significantly outperformed the baseline. For video-derived features, we identified two predictive unimodal features: learners' head posi-

tions in the x direction (p-*value* = 0.03), and the locations of their body key points in the x direction (p-*value* = 0.03). Models trained on language-based features yielded similar *MAEs* compared to the baseline model; therefore, none of the language-based features evaluated in this study were predictive for this task.

The feature space in our study is large compared to the relatively small corpus size. Therefore, identifying predictive unimodal features helped with filtering out noisy features that are not useful in predicting satisfaction scores. Next, we examined the performance of multimodal models by combining the unimodal features that were useful.

6.3 Examining Multimodal Features

For testing the performance of combining multiple features, we selected the two significant (p<0.05) unimodal features (Head Position x-axis and Body Key Points x-axis). In addition, we also selected Head Direction and Pre-trained BERT, as their *p-values* are lower than 0.1 (a threshold that has been used to identify a weak trend or association [23]). We used the best-performing unimodal model trained on Body Key Points (x-axis) as the baseline (predicted satisfaction scores $\widehat{y}_{base}$), and investigated the *p-values* of the paired-samples t-test between the predicted scores $\widehat{y}$ and the baseline results $\widehat{y}_{base}$. Table 3 shows the regression results of peer satisfaction trained on unimodal features.

The results shown in column 2 of table 3 indicated that combining Head Position (x-axis) and Head Directions yielded the lowest MAE. However, none of these multimodal features significantly improved the regression performance compared to the unimodal model.

6.4 Comparing Different Model Architectures

To understand the influence of different sequential models during feature learning, and compare the performance between linear and non-linear regression models, we selected the best-performing unimodal model and examined how prediction accuracy varied under different model architectures.

Table 4 shows the experimental results with different model architectures. For the selection of different sequential models, three models provided comparable performances, with LSTM yielding a slightly lower MAE. As for the selection of different activation functions, the model predicting satisfaction score with a *sigmoid* activation functionperformed better than with a *linear* function. In addition, although we observed faster convergence speed with *linear*, *sigmoid* provided more stable training and testing performance (see Figure 5). As for the selection of the number of layers, the one-layer LSTM performed similarly compared to two- or three-layer LSTM.

7. DISCUSSION

This study investigates the prediction of peer satisfaction using multimodal features from learners' interactions during collaborative learning activities. This section discusses the results with respect to our two research questions, as well as implications from comparing the performance of different model architectures.

Table 4: *MAEs* under different architecture settings.

Sequential Model	LSTM	RNN	GRU
MAE	0.1376	0.1401	0.1382
Output Unit	*sigmoid*	*linear*	
MAE	0.1376	0.1741	
# of Layers	1	2	3
MAE	0.1359	0.1376	0.1384

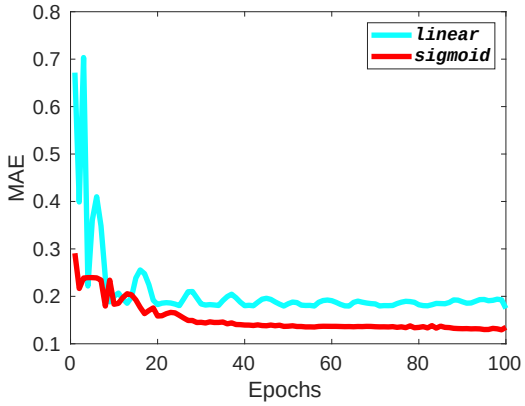

Figure 5: Testing *MAEs* under different activation functions (blue-*linear*, red-*sigmoid*). *linear* provided faster converge speed during training, while *sigmoid* provided lower MAE and more numerical stability during testing.

7.1 RQ 1: What are the most predictive unimodal features of peer satisfaction during collaboration?

7.1.1 Audio-based Features

In this study, we investigated several commonly used acoustic-prosodic features and the results showed that none of these features were significant predictors of peer satisfaction. Previous literature has associated learners' peer satisfaction with their emotional bonding [46]. Acoustic-prosodic features have been widely used for detecting speaker emotion detection (positive, neutral, and negative) [10] and predicting learners' task performance [47]. However, the results from this study indicate that the acoustic-prosodic features we tested may not have the explanatory power to predict peer satisfaction.

One potential reason for audio features not performing well in models is if there are periods of silence included in what the model thinks are periods of speech only. We examined several different silence removal thresholds (-6, -16, -30 dBFS) and the results indicated that this strategy did not help with peer satisfaction prediction. A higher silence removal threshold (e.g., -6 dBFS) could help reduce the negative influence from background noise; however, it is also more likely to remove learners' speech. While selecting between and qualitatively examining different thresholds, we determined -30 dBFS was optimal for our corpus to balance between eliminating periods of silence without excessively cutting off speech. However, acoustic features under that threshold were not predictive of peer satisfaction.

7.1.2 Language-based Features

We examined several statistical (word count and speech rate) and semantic (Word2Vec and BERT) features from the language modality. Statistical features such as word count per utterance and speech rate have shown to be associated with learners' active participation and turn-taking during collaboration [49]. The results from our study showed that there was a trend toward significance when more semantic information was added to the features (p-values for word count, Word2Vec, and BERT: 0.46, 0.16, 0.06); however, none of these models yielded statistically significant results for predicting peer satisfaction (Table 2). Previous literature also found several sequences of dialogue acts representing speakers' intentions (e.g., questions followed by clarifications) that were positively related to peer satisfaction [26], but our results did not show a direct correlation between semantics and peer satisfaction. One potential reason may be that the semantic representation methods used in our study did not have the same explanatory power as dialogue acts to directly indicate learners' intentions.

7.1.3 Video-based Features

Among the several video-based features extracted in this study, head position and body location on the horizontal axis were the only two predictive unimodal features. To better understand how the patterns of these two predictive features varied among learners with different satisfaction scores, we selected three groups of five learners and examined their sessions in more detail. The groups are as follows:

- High satisfaction group: five learners who received the highest scores (5.0 / 5.0).
- Average satisfaction group: five learners who received the exact score of 4.3 / 5.0 (mean peer satisfaction score of our corpus).
- Low satisfaction group: five learners who received the lowest five scores (all below 3.7 / 5.0).

Figure 6 shows the patterns of horizontal (x-axis) head distance from the camera, in meters, for the three groups of learners. For each group, we calculated their averaged Head Distance (x-axis) over whole sessions. Since the camera was positioned horizontally in the middle between two learners, if learners had lower head distance from the camera, this likely reflects that the learners were sitting closer to one another. From Figure 6 we could see that learners who received high satisfaction scores (green) had lower head distances over the collaborative coding sessions, compared to learners who received average (red) and low (blue) satisfaction scores.

Figure 6 also depicts the difference of the head distance variance over time across the three groups of learners. Learners who received high satisfaction scores (green) had lower head distance variance and fewer numbers of sharp distance increases over time, compared to learners who received average (red) and low (blue) satisfaction scores. A sharp head distance increase could happen when the learner became disengaged in the collaborative coding tasks (e.g., talking to learners in other groups). In comparison, for learners in the high satisfaction group (green), only a small range of head distance variance over time was observed.

In addition to head distance (x-axis), another predictive unimodal feature identified in our study was *body key points*,

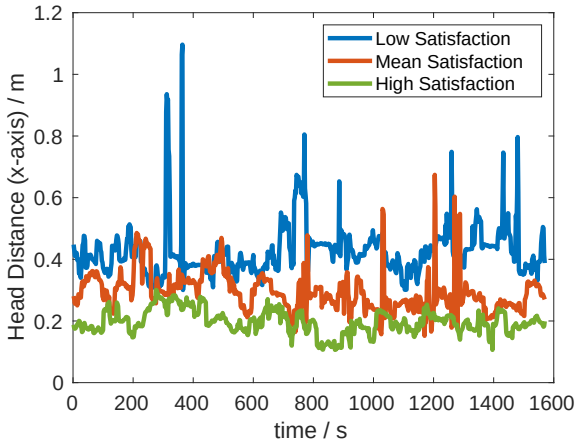

Figure 6: Head Distance (x-axis) in Meters from OpenFace

such as the location of nose, neck, and shoulders. Figure 7 shows the patterns of neck location (x-axis) in pixels toward the camera for three groups of learners; locations for other body key points followed relatively similar patterns. The maximum neck distance from the camera that could be detected was 640 pixels (half of 1280 pixels) because the resolution of our cameras was 720p. Figure 7 shows that learners who received high satisfaction scores (green) sit closer toward the camera (they had closer distances to their partners) over the collaborative coding sessions, compared to learners who received average (red) and low (blue) satisfaction scores. Additionally, learners who received high satisfaction scores (green) had lower neck location variance over time, compared to learners who received average (red) and low (blue) satisfaction scores.

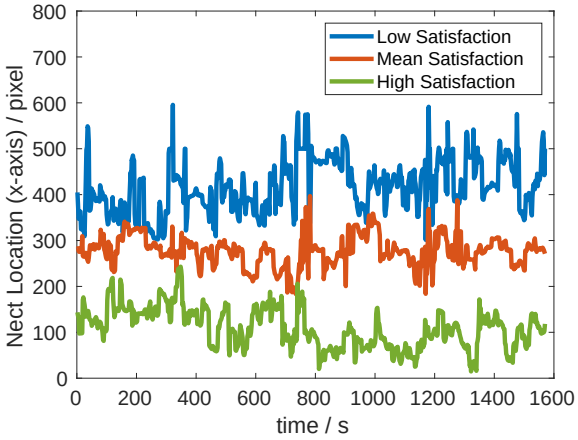

Figure 7: Neck Location (x-axis) in Pixels from OpenPose

The findings from Figure 6 and Figure 7 were aligned with previous literature that found learners' perceived social presence and proximity significantly impacted their satisfaction, as well as group performance during collaborative learning [8, 36]. For example, a study conducted by So et al. [46] revealed that learners' perceptions of physical proximity and psychological aspects of distance were both important factors in their reported satisfaction with their partner. In another similar study conducted by Spikol et al. [47], the

authors found that the distances between learners' faces and between learners' hands were two strong indicators of task performance when groups of college students were engaged in open-ended collaborative tasks.

7.2 RQ 2: Does multimodal feature fusion improve peer satisfaction prediction compared to the best-performing unimodal model?

In this study, experiment results in Table 3 from several multimodal models indicated that although using multimodal features (Head Position (x-axis) combined with Body Key Points) yielded lower *MAE* than the best-performing unimodal feature, there was no significant performance advantage of using multimodal over unimodal features. The potential reason may be that both head position and body key points represented learners' spatial locations; therefore, combining these two unimodal features did not add extra useful information to predict peer satisfaction. Therefore, each of these two unimodal features alone could be used to predict peer satisfaction. However, the head pose feature extraction process with OpenFace was faster than OpenPose. OpenPose was computationally demanding, and required GPU acceleration to perform the body key points detection. Therefore, OpenFace may be a more practical feature extraction choice over OpenPose when deploying real-time learning support systems.

7.3 Implications from comparing different model architectures

The experimental results comparing performance of different model architectures showed that the three different sequential models (RNN, LSTM, and GRU) had similar peer satisfaction prediction accuracy; in addition, non-linear regression models yielded lower MAE than linear regression models. These results have a few practical implications for researchers in the educational data mining community seeking to conduct similar studies with the methodology presented in this study.

Although sequential models were able to represent the sequential nature of utterance-level features, the comparison between different sequential models (RNN, LSTM, and GRU) did not reflect significant performance differences. Given that GRU usually has a faster training speed than LSTM and RNN due to its simpler cell structure [51], GRU could be a better choice over RNN or LSTM for similar tasks. In addition, the comparison between different activation functions (*linear* and *sigmoid*) showed that the *sigmoid* regression model yielded lower *MAE* and provided more numerical stability during testing than the *linear* model. The reason may be that the satisfaction scores predicted in this study only ranged from 1 to 5, so the constrained output value range of the *sigmoid* function could better avoid large error values during training. On the contrary, there was no mechanism to prevent the *linear* activation function from predicting out-of-range satisfaction scores.

7.4 Limitations

The current work has several important limitations. First, we only studied peer satisfaction in the context of co-located pair programming, and analyzed recordings collected from a relatively small corpus with 44 middle school learners; therefore, the predictive features found in this paper may not generalize well to group collaboration involving three or more team members, or to learners in other populations or learning environments, such as adults or online learning. Second, the LSTM-based feature learning process was black-box, which makes it relatively difficult to interpret what predictive information was learned from each unimodal feature. Third, because the satisfaction survey was administered post-hoc, after the collaboration was finished, it does not capture potential variation that may have occurred in students' attitudes toward their partners as collaboration unfolded. Finally, the effectiveness of video-derived features identified in this study relies heavily on the correct setup of the video recording process. Our dataset was collected from a natural and active classroom setting, and thus, OpenFace sometimes failed to detect both learners' faces when they were not directly facing the camera, or in the case of occlusion. Even though we used wide-angle camera lenses for video recording student interactions, there were some cases in which some students were sometimes out of the recording range.

8. CONCLUSION AND FUTURE WORK

Learners' satisfaction toward their partners plays a crucial role in group performance and learning outcomes. If intelligent systems could automatically predict peer satisfaction during collaboration, they could provide timely scaffolding for better learning experiences. In this paper, we investigated automatic prediction of peer satisfaction by analyzing 44 middle school learners' collaborative dialogues. We compared a set of state-of-the-art multimodal learning analytics techniques with linguistic, acoustic-prosodic, and visual features extracted from students' interactions. The experimental results revealed two significant predictors: head position and body location. Learners who had shorter head and body distances from their partners were more likely to receive higher peer satisfaction scores.

This study highlights several directions for future work. First, future work should examine the generalizability of the findings in this study using larger datasets, including data from online learning environments and multi-party interactions among groups of three or more learners. Second, although OpenFace and OpenPose support accurate detection of head pose and body pose, it remains challenging to integrate them into intelligent learning support systems for real-time analysis. Future work should investigate other methods and tools to detect learners' pose features accurately and time-efficiently. Finally, it is important to investigate how intelligent systems can most effectively deliver feedback to learners during collaborative learning process. As we move toward predicting peer satisfaction in real time, we will be able to build and investigate systems that can significantly improve learners' collaborative learning experience in classrooms.

9. ACKNOWLEDGMENTS

This research was supported by the National Science Foundation through grant DRL-1640141. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

10. REFERENCES

[1] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency. Openface 2.0: Facial behavior analysis toolkit. In *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 59–66. IEEE, 2018.

[2] P. Blikstein. Multimodal learning analytics. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*, pages 102–106, 2013.

[3] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[4] M. Celepkolu and K. E. Boyer. The importance of producing shared code through pair programming. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 765–770, 2018.

[5] M. Celepkolu and K. E. Boyer. Predicting student performance based on eye gaze during collaborative problem solving. In *Proceedings of the Group Interaction Frontiers in Technology*, pages 1–8. 2018.

[6] M. Celepkolu and K. E. Boyer. Thematic Analysis of Students' Reflections on Pair Programming in CS1. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education - SIGCSE '18*, pages 771–776. ACM, 2018.

[7] M. Celepkolu, D. A. Fussell, A. C. Galdo, K. E. Boyer, E. N. Wiebe, B. W. Mott, and J. C. Lester. Exploring middle school students' reflections on the infusion of cs into science classrooms. In *Proceedings of the 51st ACM technical symposium on computer science education*, pages 671–677, 2020.

[8] S. W. Chae. Perceived proximity and trust network on creative performance in virtual collaboration environment. *Procedia Computer Science*, 91:807–812, 2016.

[9] C.-L. Chan, J. J. Jiang, and G. Klein. Team task skills as a facilitator for application and development skills. *IEEE Transactions on Engineering Management*, 55(3):434–441, 2008.

[10] S. A. Chowdhury, E. A. Stepanov, G. Riccardi, et al. Predicting user satisfaction from turn-taking in spoken conversations. In *INTERSPEECH 2016*, pages 2910–2914, 2016.

[11] B. Cimatti. Definition, development, assessment of soft skills and their role for the quality of organizations and enterprises. *International Journal for Quality Research*, 10(1):97, 2016.

[12] A. D. Clarke and B. W. Tatler. Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, 102:41–51, 2014.

[13] M. Cukurova, Q. Zhou, D. Spikol, and L. Landolfi. Modelling collaborative problem-solving competence with transparent learning analytics: is video data enough? In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge*, pages 270–275, 2020.

[14] I. Daoudi, E. Tranvouez, R. Chebil, B. Espinasse, and W. Chaari. An edm-based multimodal method for assessing learners' affective states in collaborative crisis management serious games. In *Proceedings of the 13th International Conference on Educational Data Mining*, 2020.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[16] V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum. Towards collaboration translucence: Giving meaning to multimodal group data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2019.

[17] F. Eyben, M. Wöllmer, and B. Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 2010 International Conference on Multimedia*, pages 1459–1462, 2010.

[18] C. Forsyth, J. Andrews-Todd, and J. Steinberg. Are you really a team player? profiling of collaborative problem solvers in an online environment. In *Proceedings of the 13th International Conference on Educational Data Mining*. ERIC, 2020.

[19] B. Gambäck and U. K. Sikdar. Using convolutional neural networks to classify hate-speech. In *Proceedings of the 1st Workshop on Abusive Language Online*, pages 85–90, 2017.

[20] T. T. Goud, V. Smrithirekha, and G. Sangeetha. Factors influencing group member satisfaction in the software industry. In *Proceedings of the International Conference on Data Engineering and Communication Technology*, pages 223–230. Springer, 2017.

[21] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester. Automatically recognizing facial indicators of frustration: a learning-centric analysis. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 159–165. IEEE, 2013.

[22] L. Hasler-Waters and W. Napier. Building and supporting student team collaboration in the virtual classroom. *Quarterly Review of Distance Education*, 3(3):345–52, 2002.

[23] N. Houssami, P. Macaskill, M. L. Marinovich, J. M. Dixon, L. Irwig, M. E. Brennan, and L. J. Solin. Meta-analysis of the impact of surgical margins on local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy. *European Journal of Cancer*, 46(18):3219–3232, 2010.

[24] K. Huang, T. Bryant, and B. Schneider. Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data. *Proceedings of The 12th International Conference on Educational Data Mining*, 2019.

[25] E. Kapp. Improving student teamwork in a collaborative project-based course. *College Teaching*, 57(3):139–143, 2009.

[26] G. A. Katuka, R. T. Bex, M. Celepkolu, K. E. Boyer, E. Wiebe, B. Mott, and J. Lester. My partner was a good partner: Investigating the relationship between dialogue acts and satisfaction among middle school computer science learners. In *Proceedings of the 14th*

International Conference on Computer-Supported Collaborative Learning. International Society of the Learning Sciences, 2021.

[27] J. Kim, Y. Kwon, and D. Cho. Investigating factors that influence social presence and learning outcomes in distance higher education. *Computers & Education*, 57(2):1512–1520, 2011.

[28] H.-Y. Ku, H. W. Tseng, and C. Akarasriworn. Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in Human Behavior*, 29(3):922–929, 2013.

[29] Q. P. Law, H. C. So, and J. W. Chung. Effect of collaborative learning on enhancement of students' self-efficacy, social skills and knowledge towards mobile apps development. *American Journal of Educational Research*, 5(1):25–29, 2017.

[30] R. Liu, J. Davenport, and J. Stamper. Beyond log files: Using multi-modal data streams towards data-driven kc model improvement. *Proceedings of the 9th International Conference on Educational Data Mining*, 2016.

[31] C. N. Loes and E. T. Pascarella. Collaborative learning and critical thinking: Testing the link. *The Journal of Higher Education*, 88(5):726–753, 2017.

[32] M. Madaio, R. Lasko, A. Ogan, and J. Cassell. Using temporal association rule mining to predict dyadic rapport in peer tutoring. *Proceedings of the 10th International Conference on Educational Data Mining*, 2017.

[33] J. Malmberg, S. Järvelä, J. Holappa, E. Haataja, X. Huang, and A. Siipo. Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior*, 96:235–245, 2019.

[34] R. W. Mee and T. C. Chua. Regression toward the mean and the paired sample t test. *The American Statistician*, 45(1):39–42, 1991.

[35] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[36] S. Molinillo, R. Aguilar-Illescas, R. Anaya-Sánchez, and M. Vallespín-Arán. Exploring the impacts of interactions, social presence and emotional engagement on active collaborative learning in a social web-based environment. *Computers & Education*, 123:41–52, 2018.

[37] S. L. Pugh, S. K. Subburaj, A. R. Rao, A. E. Stewart, J. Andrews-Todd, and S. K. D'Mello. Say what? automatic modeling of collaborative problem solving skills from student speech in the wild. *Proceedings of The 14th International Conference on Educational Data Mining*, 2021.

[38] Pydub. `https://github.com/jiaaro/pydub`.

[39] I. Radu, E. Tu, and B. Schneider. Relationships between body postures and collaborative learning states in an augmented reality study. In *Proceedings of the 21st International Conference on Artificial Intelligence in Education*, pages 257–262. Springer, 2020.

[40] R. Rajendran, A. Kumar, K. E. Carter, D. T. Levin, and G. Biswas. Predicting learning by analyzing eye-gaze data of reading behavior. *Proceedings of the 11th International Conference on Educational Data Mining*, 2018.

[41] J. M. Reilly and B. Schneider. Predicting the quality of collaborative problem solving through linguistic analysis of discourse. *Proceedings of The 12th International Conference on Educational Data Mining*, 2019.

[42] Rev. `https://www.rev.com/`.

[43] B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. Pea. Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, 13(3):241–261, 2018.

[44] J. L. Schultz, J. R. Wilson, and K. C. Hess. Team-based classroom pedagogy reframed: The student perspective. *American Journal of Business Education*, 3(7):17–24, 2010.

[45] A. J. Sinclair and B. Schneider. Linguistic and gestural coordination: Do learners converge in collaborative dialogue?. *Proceedings of The 14th International Conference on Educational Data Mining*, 2021.

[46] H.-J. So and T. A. Brush. Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, 51(1):318–336, 2008.

[47] D. Spikol, E. Ruffaldi, L. Landolfi, and M. Cukurova. Estimation of success in collaborative learning based on multimodal learning analytics features. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pages 269–273, 2017.

[48] A. E. Stewart, Z. Keirn, and S. K. D'Mello. Multimodal modeling of collaborative problem-solving facets in triads. *User Modeling and User-Adapted Interaction*, 31(4):713–751, 2021.

[49] A. E. Stewart, Z. A. Keirn, and S. K. D'Mello. Multimodal modeling of coordination and coregulation patterns in speech rate during triadic collaborative problem solving. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 21–30, 2018.

[50] W. Wei, S. Li, S. Okada, and K. Komatani. Multimodal user satisfaction recognition for non-task oriented dialogue systems. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pages 586–594, 2021.

[51] S. Yang, X. Yu, and Y. Zhou. Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example. In *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, pages 98–101. IEEE, 2020.

[52] R. M. Zeitun, K. S. Abdulqader, and K. A. Alshare. Team satisfaction and student group performance: A cross-cultural study. *Journal of Education for Business*, 88(5):286–293, 2013.

[53] X. Zhang, Y. Meng, P. O. de Pablos, and Y. Sun. Learning analytics in collaborative learning supported by slack: From the perspective of engagement. *Computers in Human Behavior*, 92:625–633, 2019.