The Relationship between Co-Creative Dialogue and High School Learners' Satisfaction with their Collaborator in Computational Music Remixing

GLORIA ASHIYA KATUKA*, University of Florida, USA
ALEXANDER R. WEBBER, University of Florida, USA
JOSEPH B. WIGGINS, University of Florida, USA
KRISTY ELIZABETH BOYER, University of Florida, USA
BRIAN MAGERKO, Georgia Institute of Technology, USA
TOM MCKLIN, Georgia Institute of Technology, USA
JASON FREEMAN, Georgia Institute of Technology, USA

Co-creative processes between people can be characterized by rich dialogue that carries each person's ideas into the collaborative space. When people co-create an artifact that is both technical and aesthetic, their dialogue reflects the interplay between these two dimensions. However, the dialogue mechanisms that express this interplay and the extent to which they are related to outcomes, such as peer satisfaction, are not well understood. This paper reports on a study of 68 high school learner dyads' textual dialogues as they create music by writing code together in a digital learning environment for musical remixing. We report on a novel dialogue taxonomy built to capture the technical and aesthetic dimensions of learners' collaborative dialogues. We identified dialogue act n-grams (sequences of length 1, 2, or 3) that are present within the corpus and discovered five significant n-gram predictors for whether a learner felt satisfied with their partner during the collaboration. The learner was more likely to report higher satisfaction with their partner when the learner frequently acknowledges their partner, exchanges positive feedback with their partner, and their partner proposes an idea and elaborates on the idea. In contrast, the learner is more likely to report lower satisfaction with their partner when the learner frequently accepts back-to-back proposals from their partner and when the partner responds to the learner's statements with positive feedback. This work advances understanding of collaborative dialogue within co-creative domains and suggests dialogue strategies that may be helpful to foster co-creativity as learners collaborate to produce a creative artifact. The findings also suggest important areas of focus for intelligent or adaptive systems that aim to support learners during the co-creative process.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**.

Additional Key Words and Phrases: Collaboration, Co-creativity, Dialogue Acts, Computational Music, Computer Science Education

ACM Reference Format:
Gloria Ashiya Katuka, Alexander R. Webber, Joseph B. Wiggins, Kristy Elizabeth Boyer, Brian Magerko, Tom McKlin, and Jason Freeman. 2022. The Relationship between Co-Creative Dialogue and High School Learners'

Authors' addresses: Gloria Ashiya Katuka, gkatuka@ufl.edu, University of Florida, Gainesville, Florida, USA; Alexander R. Webber, University of Florida , Gainesville, USA, alex.webber@ufl.edu; Joseph B. Wiggins, University of Florida, Gainesville, Florida, USA, joseph.wiggins@ufl.edu; Kristy Elizabeth Boyer, University of Florida , Gainesville, USA; Brian Magerko, Georgia Institute of Technology, Atlanta, Georgia, USA; Tom McKlin, Georgia Institute of Technology, Atlanta, Georgia, USA; Jason Freeman, Georgia Institute of Technology, Atlanta, Georgia, USA.

2573-0142/2022/4-ART123 $15.00
https://doi.org/10.1145/3512970

Satisfaction with their Collaborator in Computational Music Remixing. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 123 (April 2022), 24 pages. https://doi.org/10.1145/3512970

1 INTRODUCTION

Co-creativity is a process by which people (or people together with machines) contribute and incorporate ideas to bring an experience or artifact into existence (e.g., [10]). The CSCW community has long studied co-creativity among collaborators in settings such as creating media content [47], coding in Minecraft [44], and co-located classroom writing [52]. All of these co-creative dialogue processes between people are characterized by *dialogue acts* that carry each person's ideas into the collaborative space through proposals, acceptance or rejection of those proposals, and feedback [6, 42].

Dialogue is a central mechanism in collaboration that allows learners to share responsibilities and actively contribute toward the same goal [53]. Recent studies in co-creativity have identified dialogue as a key contributor for connecting learners to their creative potential as they generate and share new ideas, which leads to co-constructive processes of thinking creatively together [35]. While there have been recent research efforts in understanding co-creative dialogue mechanisms, the extent to which dialogue is related to desirable outcomes is not well understood, and there is a need to further investigate dialogue in co-creative contexts [19]. In particular, while many dialogue studies have examined collaborative outcomes through post-tests or artifact analysis [53], there is also a need to investigate more affect-oriented outcomes, such as a learner's satisfaction with their collaborator during co-creative endeavors. Understanding the relationship between such outcomes and how learners interact with their partners can help us design better technologies that foster successful collaboration.

This paper reports on a study of co-creative dialogue in the context of dyads of high school learners who are co-creating computational musical artifacts in their classrooms using EarSketch, which is a web-based learning environment for coding and music remixing [15]. The learners communicated through textual messages within a synchronized coding environment for music composition and remixing (Figure 1). Each learner dyad collaborated on a creative task, usually in the form of composing an original song or remixing songs of their choice. We analyzed their textual messages (*utterances*) to identify patterns in dialogue that are significantly predictive of the learners' satisfaction with each other as collaborators. This work is guided by the following research questions:

- RQ1: What types of dialogue acts emerge during co-creative dialogue for computational music with dyads of high school learners?
- RQ2: In what ways are the dialogue acts between learners associated with self-reported peer satisfaction?

To investigate these questions, we collected a corpus of data from high school students engaged in collaborative computational music remixing; developed a dialogue taxonomy and applied it to the corpus of learners' collaborative dialogue; and identified dialogue act n-grams (sequences of length 1, 2, or 3 dialogue acts) that were significantly predictive of a learner's satisfaction with their partner. We built a model to predict learners' satisfaction with their partners using the n-grams as explanatory variables and a derived *peer satisfaction* outcome (the average of post-survey responses that captured each learner's satisfaction with their partner).

The findings reveal that n-grams generated from dialogue acts could reveal significant relationships regarding learners' satisfaction with their partner. Particularly, n-grams including *acknowledgments* and *positive feedback* were positively associated with higher peer satisfaction. On the other hand, when learner made a seemingly positive dialogue act of *accepting* following a pair of *proposals* by their partner, the n-gram including *accepting* was significantly associated with lower peer satisfaction. Further qualitative analysis of excerpts relating to these n-grams reveals important nuances in the extent to which each learner contributed to the collaborative space and how those contributions are associated with learners' satisfaction with each other. These findings inform our understanding of human-human co-creativity and hold the potential to inform the design of intelligent or adaptive systems that aim to support learners during the co-creative process.

This paper is the first to create a dialogue act taxonomy for co-creativity in computational music, and it advances our understanding of how dialogue acts are associated with peer satisfaction for high school learners collaborating in a co-creative domain. This domain differs from the many prior studies that focus on purely technical collaboration in several important ways. For example, the specifications for programming tasks in a traditional computer science context are clear and precise, with little room for learners to make creative choices about the final product. In contrast, in computational music remixing, the final product is largely in the control of the learners, and their negotiations around its aesthetics, both beforehand and while creating it, are central to the dialogue. Existing dialogue annotation schemes do not reflect these important aesthetic dialogues.

The remainder of the article is structured as follows: Section 2 reviews related work on co-creativity and collaborative coding and presents background on dialogue act analysis. Section 3 presents a brief overview of the EarSketch interface for computational music remixing. Section 4 discusses the study design, data collection, and post-survey outcomes analysis. Section 5 presents the development of the novel dialogue act taxonomy, while Section 6 presents the extraction of sub-sequences of those dialogue acts within the corpus. Section 7 describes the predictive models built upon those dialogue act sequences, and Section 8 discusses those findings with case-by-case analysis of the dialogue act patterns. Section 9 presents the conclusions and future work.

2 BACKGROUND AND RELATED WORK

The CSCW community has long studied collaboration in the context of learning, including what makes collaborative learning so effective and how it draws on numerous theories and approaches. Effective collaboration helps partners efficiently finish tasks and share knowledge, becoming assets in each other's learning [32, 53]. Partners engage in shared learning activities in a joint problem space comprising an emergent socially-negotiated set of learning goals, problem state descriptions, and problem-solving actions [21, 41]. While in these spaces, collaborative problem solving processes involve the construction of shared knowledge, negotiation/coordination, and maintaining team function [44]. Effective teams engage actively in collaborative dialogue, asking for explanations and justifications from their partners [43]. However, not all collaboration is productive. Common problems during collaborative learning include learners lack of strong collaborative skills, partners disengaging from the task, low competence, and breakdowns in the social relationship between learners and their partners [26]. Despite many research efforts in this area, there has been very little work examining how collaborative dialogue is associated with (or predictive of) a learner's satisfaction with their collaborator. For this reason, the current work focuses on investigating this area of research.

A recurrent line of research on collaboration has established the importance of collaborative dialogue between partners. A common method for exploring this dialogue is *dialogue act analysis*, which entails the review and codification of the function of utterances in a dialogue to capture the user's intention [2]. Dialogue act analysis captures the pragmatic nature of utterances, which has less to do with structure (syntax) and meaning (semantics) and more to do with context and intention (pragmatics). There is no single dialogue act analysis scheme that works for every context, since they might be unique and goal-dependent. However, dialogue act schemes are informed by decades of computational linguistics research showing how both conversational and task-oriented dialogues unfold [27, 45, 46]. While many dialogue act taxonomies exist across various domains, to the best of our knowledge, there was no existing dialogue act taxonomy that captures the dialogue processes that occur during co-creative activities. Co-creative domains in learning are distinctive because of how collaborators share and evaluate each other's ideas not merely for whether they satisfy task requirements but also along aesthetic dimensions of preference.

Among the many types of computer-supported cooperative work that occur in educational settings, collaborative coding has received increasing focus, as it has consistently been shown to create higher quality solutions and have positive social outcomes [44]. Pair programming, a common collaborative coding paradigm, can help learners persist in completing learning tasks and increase retention in college [29]. While partners are collaborating, mental model consistency helps pairs be successful in pair programming [36]. During collaboration, talking with a partner about code can also be difficult due to frequent referencing of code artifacts and the difficulty of describing intermediate code steps to achieve a goal [34]. Studies of dialogue during programming have provided insights about phenomena such as the importance of statements of uncertainty and their resolution in dialogue [40]. However, those prior studies focused on purely technical collaborations without the aesthetic component involved in computational music composition. They also did not examine the outcome of peer satisfaction as we do in the current work.

While partner satisfaction has long been recognized as an important part of collaboration, there has been little investigation on the relationship between collaborative dialogue and partner satisfaction. In the context of social partnerships within organizations, partner satisfaction is important in reflecting the extent to which collaborators are engaged or adding value within the collaboration [48]. Closer to the context of the current paper, partner satisfaction is also frequently used as an outcome in pair programming research. Pair programming studies have shown the importance of the collaborative paradigm for supporting learning, engagement, and student retention within courses [5, 11] and that students' attitudes about each other influenced the extent to which they contributed within the collaboration [7]. Additionally, peer satisfaction provides insights on how partners influence each other and how satisfied learners are when working with a partner [7, 11]. Our work utilizes peer satisfaction as an outcome metric in the context of co-creative computational music remixing, which has much in common with pair programming in that a pair of learners were contributing synchronously to a shared code artifact. Our findings advance knowledge around the co-creative interactions associated with peer satisfaction.

3 COMPUTATIONAL MUSIC IN THE EARSKETCH LEARNING ENVIRONMENT

Computational music is particularly appealing to learners as it allows them to express themselves while learning to code and create artifacts of cultural significance [15]. Learners can share their music with peers, and research indicates that these exchanges can deepen persistence in learning to code [30]. This reframing of computing to incorporate music remixing improves students' attitudes towards learning computing [14] and could be transformative for students who feel that computing is irrelevant to their lives.

The EarSketch learning environment features musical samples from various professional artists across many styles. Learners remix these sounds or make their own to create songs. EarSketch was developed to support individual learners in high school computer science classrooms, and has been used by 585,000 unique users over the past six years [15]. The EarSketch interface includes a sound browser, code editor, digital audio workstation, and a curriculum browser, which can all be accessed from the sidebar. Users can access 4,000 musical samples from the sound browser or upload their own sounds. In the code editor, users write code that algorithmically places their selected samples on a timeline. When the user runs the code, the results appear in the digital audio workstation, and users can hear the music they composed. EarSketch has several associated curricula to support learning in high school classrooms, ranging from quick one-hour experiences to a full 12-week module with pedagogical resources [15, 51].

This paper reports on work conducted in an expansion of the EarSketch interface that includes *textual chat* and synchronous code editing (Figure 1). Although there have been hundreds of thousands of users in EarSketch, this is the first study to examine the co-creative dialogue processes that unfold within it. To access the collaborative chat window, a user creates a new script and shares it (enabling edit access) with another user. By sharing an EarSketch project and exchanging synchronous textual dialogue, users can engage in co-creative processes whether the users are co-located or remote.

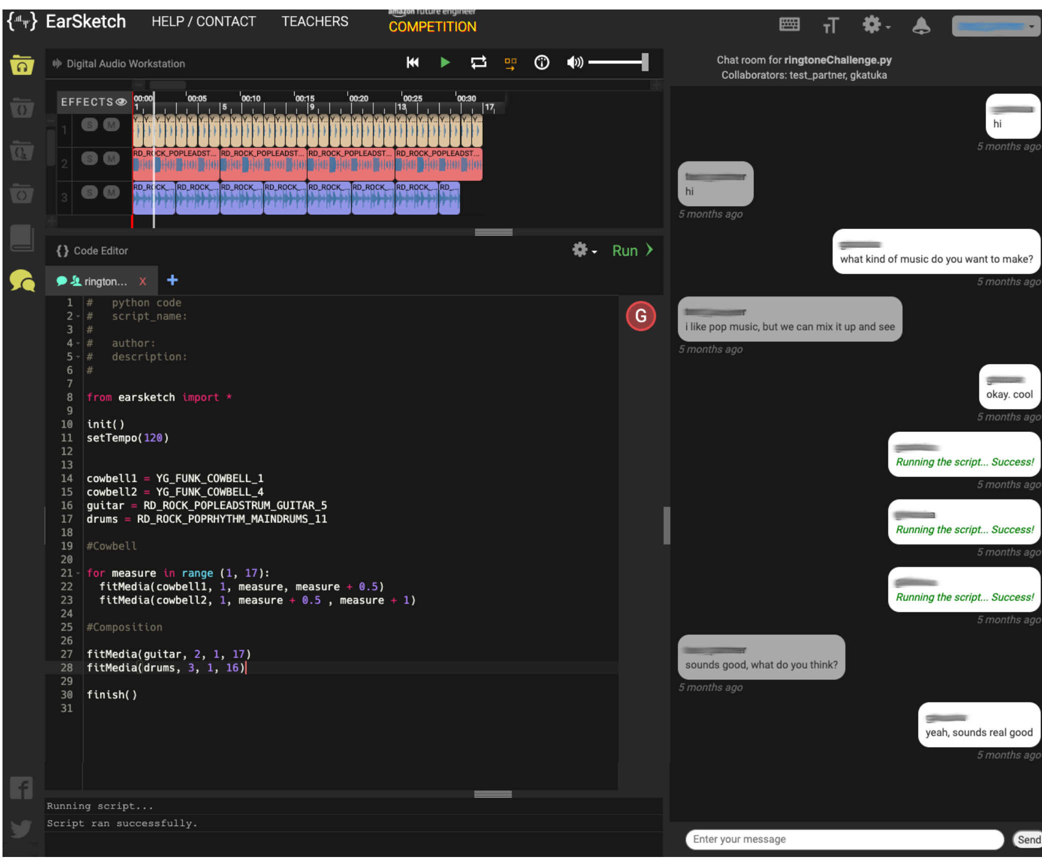


Fig. 1. The EarSketch Environment. The page shown includes the chat room with two collaborators

While much collaborative remote work today is conducted via video and voice chat, our current work focuses on textual dialogue for several reasons. First, prior work indicates that users often opt for textual chat because it is a less invasive modality that allows each collaborator time to think or access helpful resources without the social expectations of filling pauses within the dialogue [49]. Textual chat also requires less hardware and is more robust to interruptions in online connections. Finally, our work on textual dialogue for computer science collaboration over the past ten years has indicated that users often utilize the textual record offered by the chat history when they want to refer back to each other's ideas. We speculate that users may be more inclined to opt for textual dialogue while collaboratively coding in a co-creative domain than in some other collaborative domains. A convenient side effect of the textual dialogue channel is that it provides an accurate transcript of the dialogues, free from speech recognition errors and without the need for manual transcription. The study reported here takes advantage of those affordances to investigate our research questions.

4 STUDY

This study aimed to understand the co-creative dialogue patterns that emerged between physically separated dyads of high school learners using textual chat during collaborative computational music remixing. We used this data to investigate the dialogue acts that occur (Research Question 1). Then, we explored how dialogue acts between learners are associated with the outcomes reported by those learners (Research Question 2). By studying how dyads of learners work together to create music and code, our goal was to better understand how learners perceive their partners' support while collaborating.

4.1 Participants

Participants were high school learners taking either Computer Science Principles or Advanced Placement Computer Science Principles in public high schools across Georgia and Florida in the United States [1]. Between November 2019 and March 2020, 140 students from 8 schools in 2 districts in Georgia and 2 districts in Florida consented to participate in the study.[1] Among the 140 learners, 38 were from 2 classrooms in Florida and 102 from the 6 participating classrooms in Georgia. More than half of the schools were majority (>50%) Caucasian; one school was majority (>50%) Black; two schools had a substantial (between 25-35%) Latinx population; one school had a substantial (between 25-35%) Asian population. All learners were enrolled in high schools in grades 10-12 (typically between 15-18 years old) and had some prior experience with EarSketch.

4.2 Procedure

This study was conducted within high school computer science classrooms. Prior to the each classroom study the research team coordinated with the teacher to find a suitable date for learners to work on a co-creative learning task in EarSketch that could be completed within one class period. On the day of the study, researchers attended the class to facilitate the study alongside the teacher. First, learners completed a pre-survey about their experience and confidence in coding and music. Due to a logistical error, this pre-survey was only administered to one-third of the classrooms. We attempted subset analysis with the available pre-survey data, but the resulting models were weak with unstable effect sizes and are not included in this paper.

The teacher assigned learners to partners based on their usual classroom procedure (teachers are more knowledgeable on which learners are likely to work well together and which are not). Because the study aimed to examine textual dialogue, in most classes, paired learners were moved to separate sides of the classroom to promote collaboration through the interface (as opposed to through verbal dialogue). The collaboration interface allowed learner dyads to work on the same project simultaneously and communicate synchronously, as described above and shown in Figure 1. Learners worked together for an average of 48 minutes on one of two tasks. In the Ringtone task, learners were asked to co-creatively compose a 30-second ringtone. In the Cowbell task, learners

[1]Because these participants are minors, the research team first obtained parental consent by distributing consent forms through the students' teacher. Students provided assent verbally in class after a brief verbal description of the study. This data collection process was approved by the IRBs at both universities where the research team is conducting this work.

were asked to select Cowbell sounds of their choice from the sound browser and remix them by adding other sounds or modifying the order in which they played.

4.3 Measuring Outcomes: Post Survey on Partner Interactions

The learners collaborated during one class period. After their collaboration and approximately five minutes before the end of the class period, each learner completed a post survey containing seven items. The seven post-survey items were as follows: *1) "My partner helped me write better code," 2) "My partner helped me make better music," 3) "My partner valued my contribution," 4) "I enjoyed my interaction with this partner," 5) "I would like to work with this partner again," 6) "My partner made valuable contributions," and 7) "My partner helped me learn something new."* Responses were rated on a 4-point Likert scale, with 1 representing "strongly disagree," 2 representing "disagree," 3 representing "agree," and 4 representing "strongly agree." Figure 2 shows each item along with the distribution of learner responses. Researchers administered the post-survey questions to all 140 students; however, we reported on the 136 students in our finalized corpus used in subsequent analysis[2]. The vast majority of learners agreed or strongly agreed that their partner was helpful along most of the dimensions included in the post survey. This skew is likely due to several factors, one of which is the well-established tendency of learners to rate their partners highly on average, known as "leniency bias" [33]. Work on leniency bias finds that Likert responses are still valuable indicators of an individual's experience, particularly when the scale shows some variance, which is the case in the present analysis.

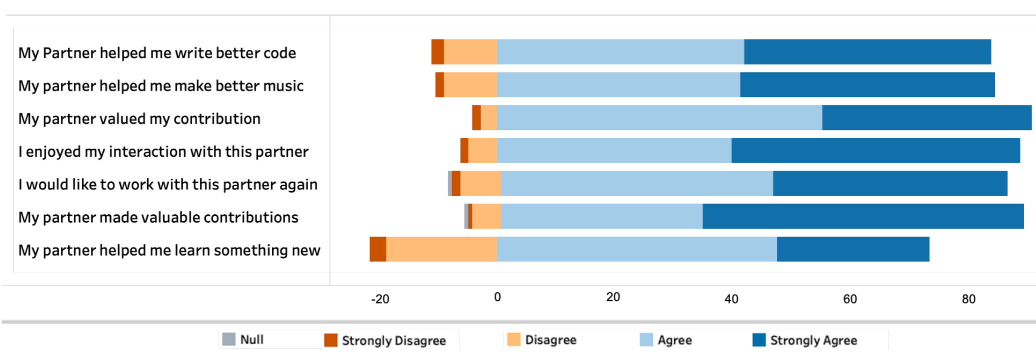


Fig. 2. Post-Survey Response on Learners' Perceptions of their Partners' Contribution

When we constructed this survey, we conceived of this co-creative space along "technical" and "aesthetic" dimensions and included corresponding survey items for "*My partner helped me write better code*" and "*My partner helped me make better music*" accordingly. However, to determine whether these items captured distinct dimensions of learners' satisfaction with their partners, we conducted a Principal Component Analysis (PCA) on all seven outcome variables collected from the post survey. The result of the PCA suggests that all seven outcome variables should be classified into one PCA cluster, which explains 62% of the variation and consists of all seven outcome members with the eigenvalue of 4.35. The Scree plot in Figure 3, which shows the eigenvalues for all possible PCA variations, with a clear drop above one component, confirms the appropriateness of one outcome dimension encompassing all seven survey items. Therefore, we averaged the post-survey outcomes for each learner to derive a single *peer satisfaction* outcome which will be treated as the dependent variable in the regression analyses reported in Table 3. On the derived

[2]See Section 4.4 for more details on data preparation that resulted in the removal of the two sessions.

peer satisfaction outcome, 75% of learners agreed or strongly agreed that they were satisfied with the overall interaction with their partner. Based on the distribution in Figure 4, the mean *peer satisfaction* rating was 3.3, with a standard deviation of 0.5 and a median of 3.4.

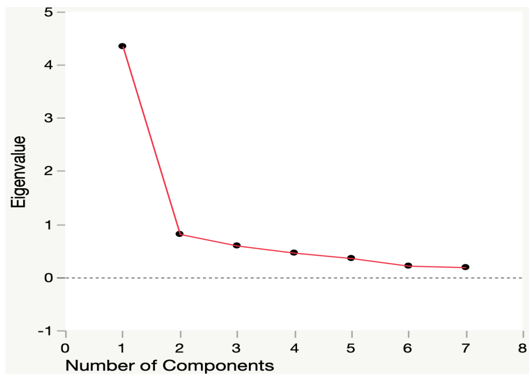


Fig. 3. Scree Plot for Post-Survey Factor Analysis

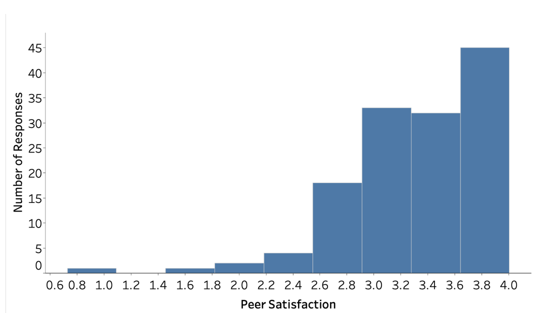
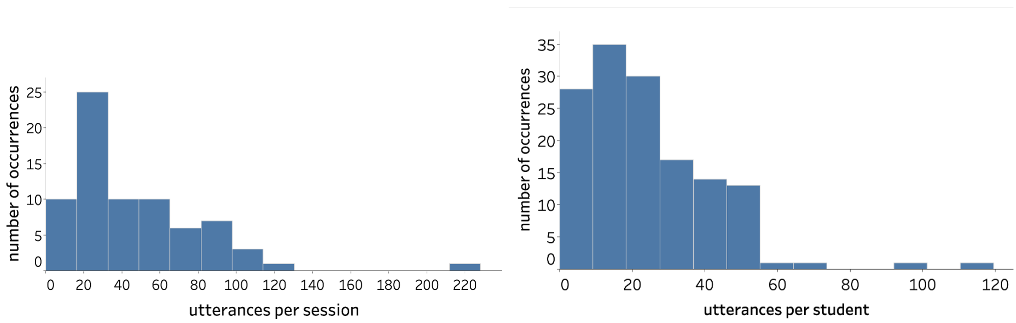


Fig. 4. Distribution of Derived Outcome Measure *Peer Satisfaction*

4.4 Dialogue Corpus

While learners interacted, we collected their textual dialogue and coding actions, which included writing code, deleting code, and executing the program. These events were written to a database in chronological order, and they produced the textual dialogue corpus we used for the present analysis. The textual dialogue corpus is made up of utterances, units of sentence-level segments of speaker turns, which are not necessarily one-to-one (a single turn can contain multiple utterances) [45]. 4533 unique chat utterances were collected over the span of several months. Researchers then excluded sessions that either: 1) included groups of three rather than pairs (in cases of an odd number of learners); 2) did not use the chat interface; or 3) engaged exclusively in off-task dialogue acts. Two sessions were removed for engaging in exclusively off-task dialogue that did not reflect that the students were engaged in the task. One of the two sessions had 121 utterances, with 20 utterances containing only the letter "s" and the remaining consisting of random letters and mostly inappropriate comments. The other session had 989 utterances, consisting of gibberish, single letters, or punctuation marks. Researchers would have annotated these utterances as Off-task

(O). This dialogue tag would not have met the threshold of occurring in 10% or more of sessions and would have been filtered out at the modeling step[3].



(a) Number of Utterances per Session

(b) Number of Utterances per Learner

Fig. 5. The Distribution of Utterances in the Corpus

The remaining dialogue corpus contains 68 sessions (136 learners) of collaborative work. Some learners worked over two days due to shorter class periods and to account for setup and technical issues. The average number of utterances per session was 48 (SD = 35, Min = 4, Max = 214), as shown in Figure 5a. For each learner, the average number of utterances was 24 (SD = 18, Min = 2, Max = 119), as shown in Figure 5b. Over the 68 sessions, 3401 utterances were collected and further analyzed.

5 DEVELOPMENT AND APPLICATION OF A TAXONOMY OF CO-CREATIVE DIALOGUE ACTS

Our goal is to analyze the utterances within these co-creative dialogues at the level of dialogue acts, as discussed in Section 2. To the best of our knowledge there is no existing taxonomy that captures the dialogue processes that occur in a co-creative learning domain. To fill this gap, we began with taxonomies that exist within closely related technical and aesthetic domains [16, 39]. First, we identified dialogue act taxonomies from collaborative coding: in particular, a closely related body of work on textual remote pair programming [39]. From that taxonomy, we included dialogue act labels such as Statement, Acknowledgement, Positive Feedback, and Non-positive Feedback. Next, we found that within improvisational theater, researchers have identified dialogue acts such as introducing a novel concept into the collaborative space, labeled as Presentation (which we term as Proposal) [16]. Other examples of dialogue acts extracted from the improvisational theater framework include directing one's collaborator to do something (adopted here as Directive), Acceptance (Proposal Acceptance and Directive Acceptance), and Rejection (Proposal Rejection and Directive Rejection) [16, 17]. Through discussion with one of the authors of the improvisational dialogue framework, we generated a set of improvisational dialogue act labels that specified how higher-order improvisational interactions might manifest in collaborative computational music remixing. The dialogue act labels created in this way include Social, Passing Responsibility, Confusion, Seeking Feedback, Closing, and Code/Link. Based on prior research on the importance of emoticons in social interactions among students in the same age group [12], we explored the presence of emoticons in our corpus. There were 39 utterances containing emoticons in the corpus. Researchers labeled the emoticons in the context in which they appeared. Twenty-eight of those

[3]See Section 6 for more details

were labeled as Social, four as Confusion, six as Passing Responsibility, and one as Seeking Feedback. An example of textual emoticons found in learners' dialogue are shown in Table 1. The entire taxonomy of co-creative dialogue acts is shown in Table 1, and the following paragraph further details the iterative process of deriving it.

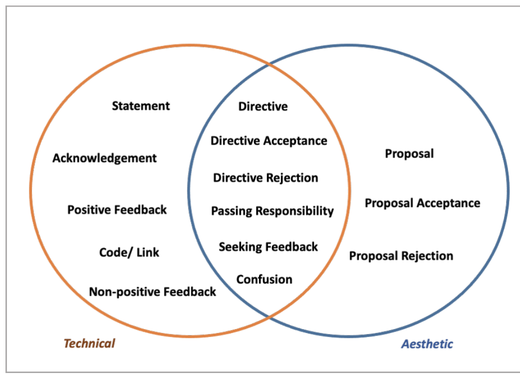


Fig. 6. The Dialogue Acts within "Technical" and "Aesthetic" Dimensions in Co-creative Dialogue

By drawing from the improvisational framework and the collaborative coding dialogue act taxonomies described above, we created an initial set of 16 tentative labels for technical and aesthetic dialogue acts, with corresponding descriptions. Examples of technical dialogue include conversations about coding elements such as functions and variables; examples of aesthetic dialogue include a learner proposing a sound, genre, or artist to use. The breakdown of the dialogue acts in terms of the technical and aesthetic dimensions captured are shown in Figure 6.

Three graduate students on the research team, whom we will refer to as annotators A, B, and C, met to apply those labels to a small sample of the corpus (approximately three dialogue sessions or 190 utterances) and identify utterances that the existing tags did not describe well. As previously mentioned in Section 4.4, the raw dataset consisted of both textual dialogue and code actions of the collaborators. The annotators referred to the code to provide context during dialogue act tagging. Additionally, annotators had access to the full history of the dialogue between the two students. I a meeting with two project leads, the graduate students presented the initially labeled data and notes on which utterances were not well captured. During the meeting, they discussed and refined the set of proposed dialogue act labels and produced updated descriptions and labels for the 16 original dialogue acts. The three graduate students applied that draft taxonomy to a new small sample of different sessions, 174 utterances, from the corpus. They computed the pairwise kappa statistic, which captures the degree of inter-annotator reliability while adjusting for the probability of agreeing by chance [25]. The pairwise kappa statistic was 0.64 between annotators A and B, 0.63 between annotators B and C, and 0.65 between annotators A and C. According to the dialogue act literature, a kappa statistic of 0.70 or greater is considered sufficiently reliable.

The lower than desired kappa statistics indicated that the labeling scheme or protocol needed further refinement. To that end, the annotators met to discuss the discrepancies between their selections and then met again with one of the project leads to refine the tagging scheme. We iterated on this refinement process a total of four times: annotators tagged a small sample separately, calculated pairwise kappas, met to review, and edited the labels and descriptions. A total of 375 utterances were used in this training process. Once sufficiently high kappa had been established, the annotators A and B proceeded to label the rest of the corpus. Annotator A labeled the entire corpus of 2851 utterances, and Annotator B labeled 918 utterances independently, resulting in a *substantial* inter-rater reliability with kappa=0.76 for the overlapping utterances. The final tagset of 16 dialogue act labels is shown in Table 1[4]

Table 1. Taxonomy of Co-creative Dialogue Acts.

Dialogue Act Label	**Rel. Freq.**	**Description**	**Examples**[4]
Statement (**STMT**)	17.14%	Utterance of information or explanation, or something that exists in the coding workspace	*we have everything* —– *We already have a fitMedia*
Social (**SOC**)	14.11%	A general salutation, off-task comment, or display of remorse that plays some social function not captured in the other tags	*hey* —– *:)*
Proposal	12.32%	An assertion of creativity, related to code or music, for the partner to consider. (When annotators felt that an utterance was borderline between **Dir** and **Proposal**, they were instructed to choose **Proposal**)	*What if we try jazz?* —– *wanna do like an ABA format?*
Directive (**DIR**)	11.56%	An utterance used to set task responsibilities for each or a single partner	*Define makeBeat on line 19* —– *you make the song i will make the code*
Confusion (**CON**)	10.41%	Seeking help, expressing confusion, lack of knowledge, or uncertainty	*Ok I messed up somewhere* —– *What is going on?*
Acknowledgement (**ACK**)	6.35%	Accepting the content of the previous utterance or series of utterances	*it'll be our chorus* ***Okay*** —– *bexause its looping the track on to itself* ***okay that makes sense***
Passing Responsibility (**PR**)	6.17%	Passing creative or technical choice to partner	*do you have a choice of what genre?*
Proposal Acceptance (**PROPOSALACCEPT**)	5.67%	(Often a response to **Proposal**) Accepting a partner's addition or assertion to the co-creative mental model shared by both partners	*what do you think we should do, jazz or dubstep* ***yeah jazzand dubstep sounds fine***
Positive Feedback (**POSFDBK**)	5.29%	Positive response relating to something the partner accomplished within the scope of the task	*[student runs code after partner fixes an error]* ***it works*** —– *[student runs code after partner edits cowbell tempo]* ***good work on the timing***
Directive Acceptance (**DIRACCEPT**)	3.97%	(Often a response to **Dir**) Response to a partner accepting the dictation of flow or direction of project	*Define makeBeat on line 19* ***ok i will figure out a makebeat***
Seeking Feedback (**SKFDBK**)	2.44%	Calling attention to or requesting comment from partner regarding one's creative contribution or state of project	*[student creates a for loop]* ***Did I do that right?***
Non-positive Feedback (**NPOSFDBK**)	2.29%	Non-positive response relating to something incorrectly done by the partner within the scope of the task	*wow those sound terrible together*
Closing (**CL**)	1.00%	Partner asserts or offers to complete the program and finish the session	*I think we should submit*
Directive Rejection (**DIRREJECT**)	0.53%	(Often a response to **Dir**) Response to a partner rejecting the dictation of flow or direction of project	*try again I was adding some stuff* ***i think its ight.***
Proposal Rejection (**PROPOSALREJECT**)	0.53%	(Often a response to **Proposal**) Rejection a partner's addition or assertion to the co-creative mental model shared by both partners	*i think i want to use yg's sounds* ***K-Pop would be a little diffculit***
Code/Link (**CDLK**)	0.21%	Code statements or snippets (Music sample titles are not necessarily code), URL or other hyperlink to material or resources (curriculum entries, links to StackOverflow)	*Applied API function makeBeat()*

Table 2 provides an excerpt from the corpus to illustrate the application of this dialogue act taxonomy. In this dyadic interaction, two learners, "Student" and "Partner", engage with each other

[4]Capitalization, punctuation, and spelling of the utterances in Table 1 *Examples*, Table 2, and Section 8 are preserved from the original students' messages, some of which contain typos.

while they co-create in EarSketch. Most conversations begin with greetings (lines 1-2), and then learners establish a common ground by the learner suggesting a sound and the partner accepting the learner's suggestion (lines 3-5). The Partner makes the next suggestion relating to the code (lines 6-11). Learners continue to build on their interaction by brainstorming what music they need to add next and how they will represent it in their code (lines 12-18). After collaboratively brainstorming, they take and pass responsibilities to each other (lines 19-23) while working on the code. They check their progress by running the code and playing the sound in the digital audio workstation. The Student confirms their progress with a positive feedback (line 25), and the Partner acknowledges and signals that they can proceed to the next step in their task (line 26). From this point on, they continue this creative process of making suggestions and confirming or rejecting each other's ideas until they complete their task.

Table 2. Excerpt 1: An Example of Co-creative Music and Code Dialogue

Utt No.	Learner	Utterance[4]	Tag
1.	Student:	hey	Soc
2.	Partner:	hey	Soc
3.	Student:	rock theme	Proposal
4.	Partner:	sounds good	ProposalAccept
5.	Student:	alright	Ack
6.	Partner:	lets start with just making a function with no arguments first	Dir
7.	Student:	k	DirAccept
8.	Partner:	you make the fitMedia	Dir
9.	Student:	ok	DirAccept
10.	Partner:	i can do the loop	Dir
11.	Student:	alright	DirAccept
12.	Student:	do we specify the music and track	Con
13.	Student:	or the start and end?	Con
14.	Partner:	we meed to do the start and end and song at least	Stmt
15.	Partner:	we can decide on track later	Proposal
16.	Student:	ok	ProposalAccept
17.	Student:	the first music can be guitar and the next can be drums or a beat	Proposal
18.	Partner:	Ok	ProposalAccept
19.	Student:	ill choose the guitar	Proposal
20.	Partner:	sounds good	ProposalAccept
21.	Student:	and you can do the beat/drums and we'll change them to fit together	Dir
22.	Partner:	yeah	DirAccept
23.	Student:	sorry	Con
24.	Student:	there	SkFdbk
25.	Student:	that sounds good	PosFdbk
26.	Partner:	yeah now lets do the next one	Dir
...	...	...	...
62.	Partner	the last guitar sounds pretty good	PosFdbk

6 EXTRACTING SEQUENCES OF DIALOGUE ACTS

Our overarching goal is to identify how dialogue acts (or sequences of them) are associated with the outcome *peer satisfaction*, as self-reported by each learner. With the labeled dialogue acts in hand, as described in the previous section, our next step was to extract subsequences of those dialogue acts, which will be treated as predictors within regression models for the outcome of *peer satisfaction*. This section describes the process of extracting those dialogue sequences.

We draw upon practices from dialogue analysis and natural language processing to extract sequences of dialogue acts using *n*-grams [27]. In natural language processing, *n*-grams refer to subsequences of length *n* of any unit of language (commonly words). Common *n* values include *n*=1 (unigrams), *n*=2 (bigrams) and *n*=3 (trigrams). *N*-grams are extracted using a "sliding window" of size *n*. For example, in the sentence, "Collaboration is very important," there are four word-level unigrams: *{(collaboration),(is),(very),(important)}*. There are three bigrams: *{(collaboration,is), (is,very), (very,important)}* and there are two trigrams: *{(collaboration,is,very), (is,very,important)}*. Although higher *n* values are sometimes valuable, they quickly become sparse and present challenges for downstream analysis. In the current work, we are particularly interested in dialogue act *n*-grams and not word-level *n*-grams. For example, in the excerpt in Table 2, the first bigram is (Soc_{stu},Soc_{par}). Lines 17, 18, and 19 of the excerpt are a trigram of dialogue acts (Proposal_{stu}, Pa_{par}, Proposal_{stu}).

We processed every dialogue from the perspective of each learner in the pair. Each individual is labeled as *stu* in one pass of their dialogue and is labeled as *par* in another pass of their dialogue. The *n*-grams, therefore, includes subscripts that indicate whether the utterance belongs to the learner (*stu*) or their partner. For example, the sequence of dialogue acts from one learner's perspective will appear as (Proposal_{stu}, Stmt_{par}), while the same interaction from their partner's perspective will appear as (Proposal_{par}, Stmt_{stu}). We extracted the *n*-grams in this way so that we would be able to predict each individual's satisfaction rating of their partner. Later, when we conduct a regression analysis, which uses the generated *n*-grams as predictors for the outcome of *peer satisfaction*, the regression will predict *stu*'s satisfaction rating of *par* [23].

This procedure generated a total of 3691 distinct *n*-grams, consisting of unigrams, bigrams and trigrams: 32 distinct unigrams (corresponding to 16 dialogue act labels, each with two possible subscripts for who said them), 646 distinct bigrams, and 3013 distinct trigrams. All bigrams and trigrams could include adjacent dialogue acts by the same student, for example, (Proposal_{stu}, Proposal_{stu}). As is common in *n*-gram analyses at any granularity (words, dialogue acts, etc.) we faced a problem of sparsity as *n* increased; some bigrams were rare, and some trigrams were very rare. It is standard to filter *n*-grams to those that meet a frequency threshold; for example, a common approach is to include only those *n*-grams that occur five or more times across the corpus. We took an intentionally inclusive approach and included all *n*-grams that occurred at least once in at least 10% of conversations. After eliminating *n*-grams that occurred in fewer than 10% of sessions, the resulting tabular data set with one row per learner (136 rows) includes 155 columns of predictors: 29 unigrams, 100 bigrams, and 26 trigrams.[5]

7 RESULTS

To explore which of the *n*-grams predicted the outcome measure of *peer satisfaction*, we conducted a regression analysis using the JMP statistical tool [20]. We entered the frequency of occurrence of each of the 155 *n*-grams for each student in a session as independent variables in a generalized regression model and used the best subset estimation method to predict *peer satisfaction*. The best subset estimation method fits all possible models at each number of predictors from zero predictors (the null model) to all *p* predictors; then, it selects the best model at each number of

[5] *n*-gram analysis scripts and sample data can be found at: https://github.com/LearnDialogue/N_gram_Gen_Dialogue_Analysis

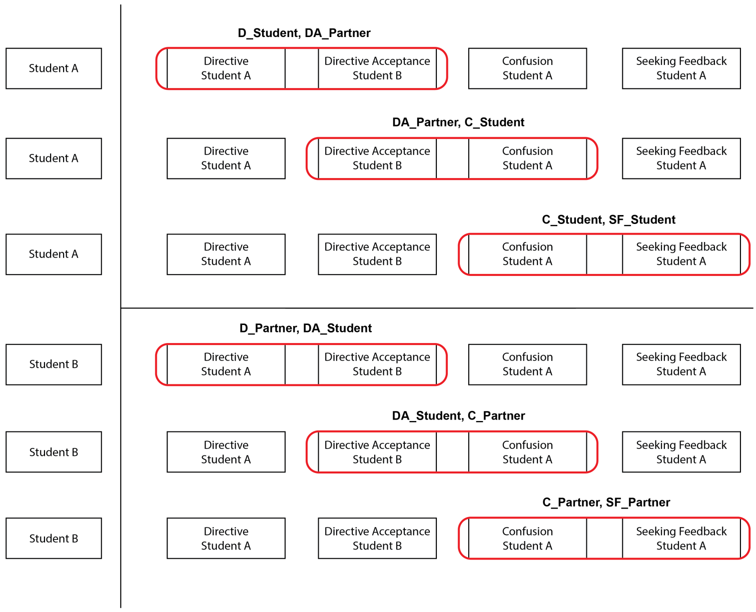


Fig. 7. Sliding Window Approach and the Bi-grams Generated from a Sequence of Dialogue

predictors, resulting in a total of p+1 best models. After this process, the algorithm compares all the best models and identifies a single best model using the Akaike Information Criterion (AIC) as a measure of goodness-of-fit. The resulting regression model identified five significant n-grams using the best subset selection method. Using the generalized regression model, we automatically adjust for multicollinearity with all the resulting variance inflation factor (VIF) values less than 1.1 (VIF values greater than five often indicate multicollinearity). We used the Benjamini-Hochberg correction method to control for false discovery rate [4]. The regression results with the adjusted p-values are shown in Table 3.

Table 3. Results of Generalized Regression using *Peer Satisfaction* as Response (n=136)

Variable	Coefficient	Std Error	p-value	Adjusted p-value
Ack_{stu}	1.7779	.0181	<.0001	.0016
PosFdbk_{par}, PosFdbk_{stu}	1.3980	.0768	.0003	.0116
Proposal_{par}, Proposal_{par}	2.0793	.0527	<.0001	.0078
Proposal_{par}, Proposal_{par}, $\text{ProposalAccept}_{stu}$	-1.7335	.1244	.0004	.0124
Stmt_{stu}, PosFdbk_{par}	-1.6183	.0910	.0002	.0103

In a dyad of students A and B, A's satisfaction is likely related to B's satisfaction because they are talking to each other and working on the same artifact. This dependence exists even though the two students completed their partner satisfaction survey independently. Additionally, we know that the dialogue acts are dependent on each other; for example, A cannot give a proposal acceptance without B making a proposal. More broadly, establishing a shared understanding entails constructing, maintaining and coordinating conversation [38, 41]. Strictly speaking, this dependence between rows in the dataset violates the assumptions of a regression model. However, we experimented with a mixed-effects model that includes a term for each pair ID, and it confirmed the same set of significant predictors presented in Table 3.

We used the parameter estimates generated by the generalized linear model with the best subset model to interpret the data, as shown in Table 3. We used the parameter estimates for centered and scaled predictors to provide standardized estimates for better interpretation (the directionality and magnitudes of the coefficients can be compared). A positive coefficient indicates that the more utterances were tagged with that given *n*-gram, the higher peer satisfaction *stu* reported. The results show that (Ack_{stu}), (PosFdbk_{par}, PosFdbk_{stu}), (Proposal_{par}, Proposal_{par}), (Proposal_{par}, Proposal_{par}, $\text{ProposalAccept}_{stu}$), and ($\text{Stmt}_{stu}$, PosFdbk_{par}) were significant predictors of the learner's satisfaction with their partner. We discuss these results in the next section.

8 DISCUSSION

We analyzed how unigrams, bigrams, and trigrams of dialogue acts are associated with a learner's rating of peer satisfaction within a corpus of co-creative dialogue for computational music remixing. The results in Section 7 revealed several co-creative dialogue act *n*-grams whose frequency was associated with a learner's satisfaction with their partner during the interaction. Three of the *n*-grams, (Ack_{stu}), (PosFdbk_{par}, PosFdbk_{stu}), and (Proposal_{par}, Proposal_{par}) were positively correlated with the *peer satisfaction* outcome. In contrast, two *n*-grams, (Proposal_{par}, Proposal_{par}, $\text{ProposalAccept}_{stu}$) and ($\text{Stmt}_{stu}$, PosFdbk_{par}) were negatively associated with the *peer satisfaction* outcome. We discuss these findings below.

8.1 Acknowledgements in Co-Creative Dialogue

As people collaborate through conversation, they establish shared understandings or plans. Dialogue theory refers to the process of establishing this "common ground" as *grounding* [8]. While many different types of dialogue acts can indicate grounding, acknowledgements and positive feedback are prime examples because they indicate that an understanding or plan has been entered into the common ground. An acknowledgment indicates that the learner understands the previous utterance made by their partner [46]. Previous studies found that utterances tagged as acknowledgment, such as *Okay*, in a collaborative dialogue provide insights about the quality of collaboration, good or bad, and about how that quality affects collaborative learning and cooperative problem solving [13, 18]. By acknowledging their partner, learners express active engagement by directly letting their partner know that they are paying attention to the conversation. As learners actively engage and accept their partner's inputs, they establish a shared understanding and common ground.

The results from our model demonstrate that in this co-creative domain, acknowledgments by the learner are positively associated with that learner's satisfaction with their partner. Responding with an acknowledgment can indicate a learner understands their partner's previous assertions. For example, one partner made the Statement, *"We need to create a custom function with 3 parameters, and call it three times, as well as creating a beat and using a for loop"* (Stmt_{par}), and the learner (*stu*) responded with the Acknowledgement, "*ok, simple enough*" (Ack_{stu}). An acknowledgment can also show that a question or confusion on the part of the learner has been remedied, providing a valuable addition to common ground from which the group can proceed [13]. For example, in one excerpt, the learner (*stu*) stated, *"Don't we have to add more fitMedias if we add more sounds,"* (Con_{stu}), and the partner responded, *"no we just have to call the function multiple times"* (Stmt_{par}), and *stu* closed with *"ok"* (Ack_{stu}).

We further examine two cases that contrast how the acknowledgment dialogue act can manifest within high-peer-satisfaction and low-peer-satisfaction dialogues. Table 4 shows example excerpts between two pairs of students, where one student (Marla[6]) reported high satisfaction (4) with her partner (Damon). In the second excerpt, the student (Sam) reported low satisfaction (1.71) with his partner (Clay) on peer satisfaction. In the first conversation, Marla and Damon collaborated to

Table 4. **Ack_{stu} (Acknowledgment by the student)**

Peer Satisfaction	Speaker Perspective	Speaker	Utterance	Tag
High (4)	Marla(student) Damon(partner)	Marla	what does the loop command do	Con
		Damon	**it makes it so that a shorter sound is repeated multiple times**	**Stmt**
		Marla	**oh okay**	**Ack**
		Marla	that should fix it i think	Stmt
		Damon	the fitmedia functions have to be fixed too	Proposal
		Damon	**its different of the powerpoint**	**Stmt**
		Marla	**ok**	**Ack**
Low (1.71)	Sam(student) Clay(partner)	Sam	yeah gimme a little	Dir
		Clay	Why'd you change my work??	Con
		Sam	sorry	SOC
		Clay	It sounds GROSSSSSSSSS	NonPosFdbk
		Clay	**I thought it sounded good, You're changing everything!!**	**Stmt**
		Sam	**im trying jeez**	**Stmt**
		Clay	This is Depressing...	NonPosFdbk

create a ringtone. In the excerpt, Marla expresses confusion about a loop command and Damon provides a helpful explanation of a loop. Marla acknowledges the explanation provided by Damon and proceeds to fix their code. Based on Marla and Damon's continued exchanges, it appears that Damon made helpful statements regularly and Marla acknowledged them. In contrast, Sam and Clay were also collaborating to create a ringtone. Sam made no acknowledgments during their session. Based on Sam and Clay's interactions, Clay provided more negative feedback and seems to have displayed an antagonistic tone. Sam's absence of acknowledgments corresponded to this less-than-positive interaction.

8.2 Positive Feedback in Co-Creative Dialogue

Feedback in this technical and aesthetic domain can include positive or negative responses towards a learner's code or musical choices. In this study, feedback refers to a response related to actions within the scope of the task. Prior studies on the role of feedback found that positive feedback may strengthen future responses, whereas negative feedback may weaken future responses [28]. In recent CSCW studies, individuals are encouraged to provide positive feedback as a way of maintaining team function and improving collaborative problem solving skills [44]. By seeking to maintain team function, learners may become more proactive contributors to the success of the collaboration.

The results of this study reinforce the importance of positive feedback by the learner, especially preceding a positive feedback from the partner. Learners reported higher satisfaction with their partners when their partners made the initial positive responses, which might have encouraged them to return positive responses as well. These positive exchanges between the learner and partner

[6] In all cases and excerpts, pseudonyms were used to maintain the anonymity of learners in our corpus.

further show a positive collaborative interaction and lead to an increase in common ground, which leaves the learner satisfied with the collaborative experience. For example, one partner said, *"well its*

Table 5. **PosFdbk_{par}, PosFdbk_{stu} (Positive Feedback by the partner followed by Positive Feedback by the student)**

Peer Satisfaction	Speaker Perspective	Speaker	Utterance	Tag
High (3.85)	Timmy(student) Simon(partner)	Simon	how do you want to make a part C?	PR
		Timmy	Oh I forgot about that	STMT
		Timmy	Maybe some precussion in the back?	PROPOSAL
		Simon	sure	PROPOSALACCEPT
		Timmy	Do you like that?	SKFDBK
		Simon	**yeah sounds good**	**POSFDBK**
		Timmy	**perfect**	**POSFDBK**
Low (2.4)	Cain(student) Kayla(partner)	Cain	How do you think CIARA_SET_DRUMBEAT_1 and CIARA_SET_DRUMBEAT_5 would sound	PROPOSAL
		Kayla	Lets find out	PROPOSALACCEPT
		Kayla	**sounds great!**	**POSFDBK**
		Cain	**i feel like it could use a little something extra**	**PROPOSAL**
		Kayla	We should add more sounds but so far so good	PROPOSAL

quite the joyful beat so far" (PosFdbk_{par}), and the learner responded with, *"i agree wholeheartidly"* (PosFdbk_{stu}). The partner utters a positive response towards how they are successfully co-creating and the learner responds with a similar positive feedback, which solidifies their bond and shows that they are in sync. Other examples of POSFDBK from our corpus are as follows: *"it sounds good"*; *"i feel like it sounds great"*; and *"There we go"* .

We further examine two cases that contrast how these dialogue acts with positive feedback can manifest within high-peer-satisfaction and low-peer-satisfaction dialogues. Table 5 shows example excerpts between two pairs of students, where one student (Timmy) reported high satisfaction (3.85) with his partner (Simon). In the second excerpt, the student (Cain) reported low satisfaction (2.4) with her partner (Kayla). In the first conversation, Timmy and Simon collaborated to remix cowbell sounds. In the excerpt, Timmy proposes a sound and seeks Simon's feedback. When Simon responds with positive feedback, Timmy reciprocates with the same sentiment. This instance of back-to-back positive exchanges may indicate that Timmy and Simon shared a strong positive connection that resulted in a positive interaction. In contrast, Cain and Kayla were also collaborating to remix cowbell sounds. However, Cain made no positive feedback responses when Kayla made a positive feedback during their session. Kayla expresses a positive sentiment towards the state of their project, but Cain did not express the same feeling and felt that their project needed more work.

8.3 Proposals in Co-Creative Dialogue

Within the current novel tagset for co-creative dialogues, a Proposal is the mechanism by which a partner introduces their ideas into the creative space in a way that their collaborator can accept, reject, or elaborate upon further (among other moves). A bigram of Proposal moves by the partner, (Proposal_{par}, Proposal_{par}), was positively associated with the learner's satisfaction. In contrast, if the bigram of Proposal moves from the partner was then followed by a Proposal Acceptance move from the learner, in a (Proposal$_{par}$, Proposal$_{par}$, ProposalAccept$_{stu}$) trigram, the learner was significantly less likely to report high peer satisfaction. Unlike the acknowledgments and positive feedback discussed in the previous subsection, acceptance of a partner's proposals may reveal important nuances in the way common ground and shared plans are established in co-creative dialogue.

A pair of Proposals from the partner is positively associated with the learner's rating of that partner. A second proposal move adjacent to the first can indicate elaborating on an idea, such as *"lets add some sort of build up"* (Proposal$_{par}$), followed by *"at 10 seconds it should all groove together"* (Proposal$_{par}$). A second Proposal move can also provide an additional alternative for the learner to explore, for example, *"Should we have Guitar?"* (Proposal$_{par}$), followed by, *"and Piano?"* (Proposal$_{par}$). Importantly, this pair of Proposal moves by the partner can make way for the learner to respond in many different ways, including by accepting the proposal (15 occurrences in the corpus); making their own proposal (8 occurrences in the corpus) and issuing a directive (5 occurrences in the corpus). Among these possibilities, a learner accepting their partner's proposal is negatively associated with that learner's rating of their partner. We believe this may be because simply accepting a proposal is among the more content-free ways to reply to that proposal, and may indicate the learner is experiencing slightly less control or investment. For example, one partner said, *"dubstep edm, drum or bass"* (Proposal$_{par}$) followed by *"was thinking just some basic dubstep"* (Proposal$_{par}$), and the learner responded with, *"ok sure"*(ProposalAccept$_{par}$). It is possible that in cases like this one, the bigram of proposal moves by the partner did not leave space for the learner to contribute further contentful ideas; that social expectations may have influenced the learner's acceptance of the proposals; or that the learner was slightly less invested than at other times in the dialogue.

We further examine two cases that contrast how the proposal dialogue act can manifest within high-peer-satisfaction and low-peer-satisfaction dialogues. Table 6 shows excerpts between two pairs of students, where one student, Drake, reported high satisfaction (3.71) with his partner Melia. In the second excerpt, the student (Stella) reported low satisfaction (2.2) with her partner Frank. In the first conversation, Drake and Melia collaborated to create a ringtone. In the excerpt, Melia makes back-to-back proposals, which led to Drake also proposing an additional idea. These creative exchanges may indicate that Drake felt comfortable expressing his ideas and building on Melia's ideas. In contrast, Stella and Frank were collaborating to remix cowbell sounds. Frank made no back-to-back proposals during their session. Based on Stella and Frank's interactions, Stella seems to struggle to understand Frank's proposed idea without an additional explanation, and passes the responsibility of implementing the idea to Frank. These instances may suggest that Frank and Stella were not establishing shared ideas or common ground, and they needed to instruct each other on what to do next.

We now shift our attention to an n-gram that was correlated with lower peer satisfaction. Table 7 shows example excerpts between two pairs of students, where one student (Sarah) reported low satisfaction (2.5) with her partner Reese. In the second excerpt, the student (Bailey) reported high satisfaction (3.85) with her partner (Jay). In the first conversation, Sarah and Reese collaborated to create a ringtone. In the excerpt, Reese makes two separate proposals, and Sarah accepts the proposal but does not seem to fully like the suggestions. Based on Sarah and Reese's interactions, Sarah accepted Reese's proposal with some reservations pending whether or not it worked. In contrast, Bailey and Jay were also collaborating to create a ringtone. In the excerpt, Jay makes back-to-back proposals, and Bailey also makes a proposal. This exchange may indicate that Bailey had a clear understanding of Jay's proposals and was comfortable enough to express her own ideas.

Table 6. **Proposal$_{par}$, Proposal$_{par}$ (Proposals by the partner)**

Peer Satisfaction	Speaker Perspective	Speaker	Utterance	Tag
High (3.71)	Drake(student) Melia(partner)	Melia	alright we'll have our sounds based around that	Proposal
		Drake	ok	ProposalAccept
		Melia	**lets add some sort of build up**	**Proposal**
		Miriam	**at 10 seconds it should all groove together**	**Proposal**
		Drake	we need sum upbeat noice sounding noises	Proposal
Low (2.2)	Stella(student) Frank(partner)	Stella	which one do you want to do?	PR
		Frank	**lets do a normal song**	**Proposal**
		Stella	**do you want to start**	**PR**
		Frank	i think u should start it	Dir
		Stella	do you want to start putting in the music	PR
		Frank	sure	DirAccept

These possibilities point toward important future investigations within co-creative dialogues for learning.

Table 7. Proposal$_{par}$, Proposal$_{par}$, ProposalAccept$_{stu}$ **(Proposals by the partner followed by a Proposal Acceptance by the student)**

Peer Satisfaction	Speaker Perspective	Speaker	Utterance	Tag
Low (2.57)	Sarah(student) Reese(partner)	Reese	**Should we have Guitar?**	**Proposal**
		Reese	**and Piano?**	**Proposal**
		Sarah	**maybe we gotta make it cool**	**ProposalAccept**
		Sarah	add it we will see how it sounds	Dir
		Reese	I'm not quite sure make it cool gonna work	Con
High (3.85)	Bailey(student) Jay(partner)	Jay	**lets set an effect**	**Proposal**
		Jay	**on our music**	**Proposal**
		Bailey	**we can increase the volume of yours because its kinda quiet**	**Proposal**
		Jay	bruh	Soc
		Bailey	bruh who	Soc
		Bailey	that was youuuu	Soc

8.4 Statements followed by Positive Feedback in Co-Creative Dialogue

Similarly, a Statement by the learner (*stu*) followed by positive feedback by their partner (*par*) is associated with lower satisfaction reported by the learner. In our corpus, one of the most common dialogue acts is an utterance of information or explanation, which is usually tagged as a Statement. Previous research has found that the Statement tag, which may also be an assertion that states or provides information about something the speaker is doing, can co-occur with most of the other dialogue acts [24, 50]. This finding is reflected in our dataset of generated *n*-grams, where the Statement tag co-occurs with almost every other dialogue act. Of the dialogue acts that co-occur with Statements, our results show that a statement by the learner followed by positive feedback by their partner was found to be correlated with a less positive collaborative experience for the learner. We believe this might have a similar explanation as the dynamic surrounding Proposals (Section 8.3), where the other possible responses a partner could have undertaken may have been preferable to simply providing positive feedback on the learner's statement.

For example, one learner said, *"I'm making megalovania"* (STMT$_{stu}$) and their partner replied, *"great"* (POSFDBK$_{par}$); or, another learner said, *"I know, we aient done yet."* (STMT$_{stu}$), to which their partner replied, *"yeah but it does sound nice"* (POSFDBK$_{par}$); and one *stu* said, *"we needed to have start measure and end measure inside the function"* (STMT$_{stu}$), to which *par* replied, *"works now gucii"* (POSFDBK$_{par}$). In all of these cases, it may be that the partner was displaying slightly less investment in the shared task (providing relatively lower content within their utterances than the student's statement may have afforded), and the higher frequency of this behavior was associated with the learner rating their partner lower. It is also possible that this pair of dialogue acts can represent indirect disagreements on the aesthetics of the artifact; for example, one learner said, *"sounds like mario kart wii"* (STMT$_{stu}$), and their partner replied, *"but i like it"* (POSFDBK$_{par}$).

We further examine two cases that contrast how the statement and positive feedback dialogue acts can manifest within high-peer-satisfaction and low-peer-satisfaction dialogues. Table 8 shows example excerpts between two pairs of students, where one student (Ryan) rated his partner (Kevin) low on peer satisfaction (2). In the second excerpt, the student (Steve) rated his partner (Mark) high on peer satisfaction (3.85). In the first conversation, Ryan and Kevin collaborated to create a ringtone. Ryan made the highest number of statements followed by a partner's positive feedback in the corpus, and he rated Kevin a 2 on peer satisfaction. In the excerpt, Ryan makes a statement and Kevin responds with positive feedback, then goes on to seek feedback on his own assertions. From Ryan and Kevin's exchanges, Kevin was more focused on getting feedback on his work than acknowledging Ryan's inputs. In contrast, Steve and Mark were collaborating to remix cowbell sounds. Mark made no positive feedback response to any of Steve's statements during their session and Steve rated Mark a 3.85 on peer satisfaction. Based on Steve and Mark's interactions, Mark acknowledged Steve's statements before making a positive feedback. Mark's acknowledgment may indicate that Mark is paying attention to Steve and may have corresponded to the positive interaction that the overall model revealed was strongly predictive of peer satisfaction.

Table 8. STATEMENT$_{stu}$, POSFDBK$_{par}$ **(Statement by the student followed by Positive Feedback by the partner)**

Peer Satisfaction	Speaker Perspective	Speaker	Utterance	Tag
Low (2)	Ryan(student) Kevin(partner)	Ryan	**because you're a doop, it tells you what's wrong with the script at the bottom of the page**	**STMT**
		Kevin	**yeah i like it, thats a good start**	**POSFDBK**
		Kevin	what do you think of these drums	SKFDBK
		Kevin	are they to rock	SKFDBK
		Ryan	it sounds like a video game ngl	STMT
High (3.85)	Steve(student) Mark(partner)	Mark	we are winging it	PROPOSAL
		Steve	**i really wanna listen to it lol**	**STMT**
		Mark	**yeah**	**ACK**
		Mark	**it sounds good**	**POSFDBK**
		Steve	i feel like it sounds great	POSFDBK

8.5 Design Considerations

Our work suggests important design considerations for future intelligent or adaptive systems that aim to support learners during the co-creative process. In such systems, a co-creative agent could collaborate with a learner during co-creative activities or tasks instead of a human partner. For example, the agent could be designed with naturalistic co-creative dialogue that is responsive to human co-creators [37], which could improve the user experience with the co-creative system [22]. Our study examines the preferred dialogue exchanges of learners by examining which dialogue patterns were associated with peer satisfaction. Designers of co-creative systems should consider avoiding dialogue patterns that have been shown to be negatively associated with partner satisfaction in human-human studies. For instance, to avoid the potentially unhelpful pattern we found of two adjacent proposals by the partner (agent) followed by acceptance by the learner, a co-creative agent would detect when learners are repeatedly accepting agent proposals without adding their own ideas and make adjustments to the dialogue to encourage learners to make their own proposals. Encouraging learners to make their own proposals is also emphasized in work on human-computer co-creativity [9]. In a complementary vein, agents could be designed to support positive dialogue patterns, such as affording acknowledgments by the students, and the ability for the student to provide positive feedback in response to the partner's (agent's) positive feedback. These design recommendations can help move the field of co-creative agents toward becoming partners who actively participate in the process of a conversation, rather than just providing responses.

8.6 Limitations

As one of the first studies to focus on co-creative dialogue within a technical and aesthetic domain, this work begins to contribute an understanding of how learners collaborate to achieve a shared creative goal. Because of the importance of social relationships between partners in this study, it is likely that learners' familiarity with, and perception of, their partner prior to interaction may have influenced their resulting satisfaction. We did not collect data on the pre-existing relationship between learners, and future work should investigate this open question since previous studies have shown that familiarity between partners can influence collaboration by increasing the levels of reasoning, quality of work, and comfort levels between partners [3, 31]. Additionally, due to the logistical error in collecting the pre-survey data, we were unable to investigate the effect of the pre-survey items on learners' satisfaction with their partner. Although we attempted subset analysis on the available pre-survey data, the resulting models were weak with unstable effect sizes. However, the results suggest important directions for future work. Also, this study was conducted in the context of high schools in the United States, and further studies are needed to examine whether the findings generalize beyond that population of learners. Furthermore, the study was conducted in textual dialogue for the reasons described in Section 4, and the findings are likely to differ when examining spoken dialogue because of its higher bandwidth requirements and different turn-taking rules (for example, speech overlap manifests as sequential contributions in textual dialogue rather than as overlap). The research questions posed here are important to investigate in the context of spoken dialogue as well. Finally, the study did not include controlled experimental conditions and therefore cannot establish causality. That is, the findings demonstrate which *n*-grams are *associated* with higher peer satisfaction, but do not demonstrate that those *n*-grams *caused* the higher satisfaction. Investigating additional factors that are influential within this complex co-creative process is an important future step.

9 CONCLUSION AND FUTURE WORK

The complexity of building and maintaining a co-creative partnership makes collaborative dialogues for co-creativity an important area of research. While many years of research have begun to shed light on the mechanisms by which collaborative dialogue is effective, co-creative dialogue for learning is understudied. Additionally, very few studies prior to this one have modeled peer satisfaction as an outcome metric of collaborative dialogues. We have examined high school learners' perception of their partner in a coding and music co-creative domain. The results show a significant relationship between co-creative dialogue features and peer satisfaction. Specifically, the findings highlight the importance of acknowledgments and positive feedback for establishing common ground, while providing ideas (proposals) in a way that fosters engagement and investment for both learners.

Future work should examine dialogue patterns that emerge within textual or spoken dialogue between different populations of learners using collaborative dialogue analysis, and should study the impact of partners' contributions to the collaborative processes. For example, with a larger sample size, further research into the individual post-survey items which we aggregated here could reveal additional insights: for instance, with a sufficiently large sample, learners' responses to the specific items "My partner helped me write better code" and "My partner helped me make better music" could be used to more clearly assess whether the partner's contributions were more technical or aesthetic during the collaboration. Another important extension of this work involves the nature of the tasks performed by the students. It is likely that the nature of the task has an influence on the co-creative dialogue. For example, the ringtone task examined in this paper may have been more personally relevant than the cowbell task. The influence of task on co-creative dialogue is an important consideration for future work. In addition, these dialogue patterns can identify valuable insights for designing co-creative artificial intelligence that could enhance human collaborative work. The current work examined textual dialogue, and in addition to the textual dialogue considered here, other modalities including speech, gaze, and nonverbal communication within both remote and co-located co-creativity should be investigated. Through these lines of investigation, we can advance understanding of human co-creativity and how to foster it with technology.

ACKNOWLEDGMENTS

Thanks to the members of the LearnDialogue Group for their help, as well as the reviewers for their valuable suggestions and suggestions for improvement. This material is based upon work supported by the National Science Foundation under grant DRL-1814083. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

REFERENCES

[1] Owen Astrachan and Amy Briggs. 2012. The CS principles project. *ACM Inroads* 3, 2 (2012), 38–42.

[2] John Langshaw Austin. 1975. *How to do things with words*. Vol. 88. Oxford university press.

[3] Margarita Azmitia and Ryan Montgomery. 1993. Friendship, transactive dialogues, and the development of scientific reasoning. *Social development* 2, 3 (1993), 202–221.

[4] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[5] Jennifer Bevan, Linda Werner, and Charlie McDowell. 2002. Guidelines for the use of pair programming in a freshman programming class. In *Proceedings 15th Conference on Software Engineering Education and Training (CSEE&T 2002)*. IEEE, 100–107.

[6] Harry Bunt. 2005. A framework for dialogue act specification. *Proceedings of SIGSEM WG on Representation of Multimodal Semantic Information* (2005).

[7] Shannon Campe, Jill Denner, Emily Green, and David Torres. 2020. Pair programming in middle school: variations in interactions and behaviors. *Computer Science Education* 30, 1 (2020), 22–46.

[8] Herbert H Clark and Edward F Schaefer. 1989. Contributing to discourse. *Cognitive science* 13, 2 (1989), 259–294.

[9] Nicholas Davis. 2013. Human-computer co-creativity: Blending human and computational creativity. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 9.

[10] Nicholas Davis, Chih-Pin Hsiao, Yanna Popova, and Brian Magerko. 2015. An enactive model of creativity for computational collaboration and co-creation. In *Creativity in the Digital Age*. Springer, 109–133.

[11] Jill Denner, Linda Werner, Shannon Campe, and Eloy Ortiz. 2014. Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education* 46, 3 (2014), 277–296.

[12] Daantje Derks, Arjan ER Bos, and Jasper Von Grumbkow. 2007. Emoticons and social interaction on the Internet: the importance of social context. *Computers in human behavior* 23, 1 (2007), 842–849.

[13] Sidney D'Mello, Andrew Olney, and Natalie Person. 2010. Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining* 2, 1 (2010), 1–37.

[14] Shelly Engelman, Brian Magerko, Tom McKlin, Morgan Miller, Doug Edwards, and Jason Freeman. 2017. Creativity in authentic STEAM education with EarSketch. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 183–188.

[15] Jason Freeman, Brian Magerko, Doug Edwards, Tom Mcklin, Taneisha Lee, and Roxanne Moore. 2019. EarSketch: engaging broad populations in computing through music. *Commun. ACM* 62, 9 (2019), 78–85.

[16] Daniel Fuller and Brian Magerko. 2010. Shared mental models in improvisational performance. In *Proceedings of the Intelligent Narrative Technologies III Workshop*. 1–6.

[17] Daniel Fuller and Brian Magerko. 2011. Shared mental models in improvisational theatre. In *Proceedings of the 8th ACM conference on Creativity and cognition*. 269–278.

[18] Bradley A Goodman, Frank N Linton, Robert D Gaimari, Janet M Hitzeman, Helen J Ross, and Guido Zarrella. 2005. Using dialogue features to predict trouble during collaborative learning. *User Modeling and User-Adapted Interaction* 15, 1-2 (2005), 85–134.

[19] Sara Hennessy, Sylvia Rojas-Drummond, Rupert Higham, Ana María Márquez, Fiona Maine, Rosa María Ríos, Rocío García-Carrión, Omar Torreblanca, and María José Barrera. 2016. Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction* 9 (2016), 16–44.

[20] SAS Institute Inc. 2021. *JMP® Pro 15*. https://www.jmp.com/en_us/software/predictive-analytics-software.html

[21] Heisawn Jeong and Michelene TH Chi. 1997. Construction of shared knowledge during collaborative learning. *Computer Support for Collaborative Learning'97 Proceedings* (1997), 124–128.

[22] Anna Aurora Kantosalo, Jukka Mikael Toivanen, Hannu Tauno Tapani Toivonen, et al. 2015. Interaction evaluation for human-computer co-creativity: A case study. In *Proceedings of the sixth international conference on computational creativity*. Brigham Young University.

[23] Gloria Ashiya Katuka, Richard T Bex, Mehmet Celepkolu, Kristy Elizabeth Boyer, Eric Wiebe, Bradford Mott, and James Lester. 2021. 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-CSCL 2021*. International Society of the Learning Sciences.

[24] Su Nam Kim, Lawrence Cavedon, and Timothy Baldwin. 2010. Classifying dialogue acts in one-on-one live chats. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. 862–871.

[25] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[26] Ha Le, Jeroen Janssen, and Theo Wubbels. 2018. Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education* 48, 1 (2018), 103–122.

[27] Diane Litman and Kate Forbes-Riley. 2006. Correlations between dialogue acts and learning in spoken tutoring dialogues. *Natural Language Engineering* 12, 2 (2006), 161.

[28] Gloria Mark. 2002. Conventions and commitments in distributed CSCW groups. *Computer Supported Cooperative Work (CSCW)* 11, 3-4 (2002), 349–387.

[29] Charlie McDowell, Linda Werner, Heather E Bullock, and Julian Fernald. 2006. Pair programming improves student retention, confidence, and program quality. *Commun. ACM* 49, 8 (2006), 90–95.

[30] Tom McKlin, Brian Magerko, Taneisha Lee, Dana Wanzer, Doug Edwards, and Jason Freeman. 2018. Authenticity and personal creativity: How EarSketch affects student persistence. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 987–992.

[31] Dorothy Miell and Raymond MacDonald. 2000. Children's creative collaborations: The importance of friendship when working together on a musical composition. *Social Development* 9, 3 (2000), 348–369.

[32] Robert C Miller, Haoqi Zhang, Eric Gilbert, and Elizabeth Gerber. 2014. Pair research: matching people for collaboration, learning, and productivity. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social*

computing. 1043–1048.

[33] Janet Morahan-Martin. 1996. Should peers' evaluations be used in class projects?: Questions regarding reliability, leniency, and acceptance. *Psychological Reports* 78, 3_suppl (1996), 1243–1250.

[34] Steve Oney, Christopher Brooks, and Paul Resnick. 2018. Creating guided code explanations with chat. codes. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

[35] Manoli Pifarre. 2019. Using interactive technologies to promote a dialogic space for creating collaboratively: A study in secondary education. *Thinking Skills and Creativity* 32 (2019), 1–16.

[36] Alex Radermacher, Gursimran Walia, and Richard Rummelt. 2012. Improving student learning outcomes with pair programming. In *Proceedings of the ninth annual international conference on International computing education research*. 87–92.

[37] Prashanth Thattai Ravikumar, Kevin McGee, and Lonce Wyse. 2018. Back to the experiences: empirically grounding the development of musical Co-creative partners in Co-experiences. In *6th International Workshop on Musical Metacreation. 9th International Conference on Computational Creativity, ICCC*. 1–7.

[38] Daniel C Richardson, Rick Dale, and Natasha Z Kirkham. 2007. The art of conversation is coordination. *Psychological science* 18, 5 (2007), 407–413.

[39] Fernando J. Rodríguez, Kimberly Michelle Price, and Kristy Elizabeth Boyer. 2017. Exploring the pair programming process: Characteristics of effective collaboration. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 507–512.

[40] Fernando J Rodríguez, Kimberly Michelle Price, and Kristy Elizabeth Boyer. 2017. Expressing and addressing Uncertainty: A study of collaborative Problem-solving dialogues. Philadelphia, PA: International Society of the Learning Sciences.

[41] Jeremy Roschelle and Stephanie D Teasley. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*. Springer, 69–97.

[42] Alexander Schmoelz. 2017. On co-creativity in playful classroom activities. *Creativity. Theories–Research-Applications* 4, 1 (2017), 25–64.

[43] Amy Soller, Alan Lesgold, Frank Linton, and Brad Goodman. 1999. What makes peer interaction effective? Modeling effective communication in an intelligent CSCL. In *Proceedings of the 1999 AAAI fall symposium: Psychological models of communication in collaborative systems*. Cape Cod, 116–123.

[44] Angela EB Stewart, Hana Vrzakova, Chen Sun, Jade Yonehiro, Cathlyn Adele Stone, Nicholas D Duran, Valerie Shute, and Sidney K D'Mello. 2019. I say, you say, we say: using spoken language to model socio-cognitive processes during computer-supported collaborative problem solving. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.

[45] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics* 26, 3 (2000), 339–373.

[46] David R Traum and Elizabeth A Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational intelligence* 8, 3 (1992), 575–599.

[47] Anna Vasilchenko, Ghaith Tarawneh, Haneen Qarabash, and Madeline Balaam. 2018. Collaborative Content Creation: Impact of Media Type on Author Behavior. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 341–344.

[48] Sandra A Waddock and Brendan D Bannister. 1991. Correlates of effectiveness and partner satisfaction in social partnerships. *Journal of Organizational Change Management* (1991).

[49] April Yi Wang, Anant Mittal, Christopher Brooks, and Steve Oney. 2019. How data scientists use computational notebooks for real-time collaboration. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.

[50] Tianhao Wu, Faisal M Khan, Todd A Fisher, Lori A Shuler, and William M Pottenger. 2005. Posting act tagging using transformation-based learning. In *Foundations of data mining and knowledge discovery*. Springer, 319–331.

[51] Anna Xambó, Jason Freeman, Brian Magerko, and Pratik Shah. 2016. Challenges and new directions for collaborative live coding in the classroom. In *International Conference of Live Interfaces (ICLI 2016). Brighton, UK*.

[52] S Yim, D Wang, JS Olson, V Vu, and M Warschauer. 2017. Synchronous writing in the classroom: Undergraduates' collaborative practices and their impact on text quality, quantity, and style. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'17)*, Vol. 10.

[53] Gang Zeng and Shigenobu Takatsuka. 2009. Text-based peer–peer collaborative dialogue in a computer-mediated learning environment in the EFL context. *System* 37, 3 (2009), 434–446.

Received January 2021; revised July 2021; accepted November 2021