International Journal of Engineering Education Vol. 40, No. 5, pp. 998–1012, 2024
Printed in Great Britain
0949-149X/91 $3.00+0.00

Harnessing Generative AI to Enhance Feedback Quality in Peer Evaluations within Project-Based Learning Contexts*

SUSAN SAJADI, MARK HUERTA, OLIVIA RYAN and KATIE DRINKWATER
Virginia Tech, Department of Engineering Education, 375 Goodwin Hall, Blacksburg, VA 24060, USA.
E-mail: ssajadi@vt.edu, markhuerta@vt.edu, oryan222@vt.edu, kdd20@vt.edu

This paper explores using generative artificial intelligence (AI) large language models (LLMs) to augment traditional peer feedback processes within project-based learning (PBL) environments where students work in teams. Our study contributes to the expanding digital educational frontier by harnessing the power of LLMs for summarizing and enhancing self and peer feedback. We present a novel educational innovation that utilizes GPT-4 to effectively summarize self and peer comments for student teams in PBL courses. Our research involved the development of a specific prompt grounded in relevant educational theories and the analysis of 118 AI-summarized Performance Feedback Reports that were generated for students. The findings reveal that the LLM significantly improves the quality of peer feedback by making it more constructive and actionable, addresses off-topic or inappropriate comments, and provides a meaningful comparison between a student's self-evaluation and peer comments. Our paper underscores the importance of peer feedback systems as valuable pedagogical tools in PBL classes and demonstrates how AI tools can enhance these systems. The outcomes of this research serve as a foundational guide for faculty, especially those teaching in PBL environments with student teams, who are considering using AI to enhance teaching and learning. We demonstrate the potential of integrating these innovative practices in PBL classrooms to support students' development of teamwork skills, marking a significant step towards a new digital frontier in engineering education.

Keywords: generative AI; LLMs; feedback; teamwork; problem or project-based learning (PBL)

1. Introduction

Developing and nurturing engineering students' teamwork skills is vital to their professional formation, substantiated by its inclusion in the Accreditation Board for Engineering and Technology (ABET) student outcome criteria. Engineering work is characteristically collaborative, with teamwork recognized as an essential competency alongside other professional skills such as problem-solving and communication [1]. Project-based learning (PBL) courses often offer a platform for students to engage in team-oriented engineering design projects that help foster professional skills. In PBL contexts, peer evaluation systems such as CATME (Comprehensive Assessment for Team Member Effectiveness), ITP (Individual Team Performance) Metrics, and TEAMMATES have emerged as crucial pedagogical tools for facilitating teaming. These tools assist engineering instructors in forming teams, monitoring team dynamics, and providing formative and summative peer feedback to students on their teamwork behaviors and individual performance. Research supports peer evaluations as a tool that can promote student accountability and help mitigate social loafing [2–5], the tendency of some students to exert less effort than their peers when working in a group.

While peer evaluation systems have been used as instructional tools for monitoring team dynamics over the years, their current design limits their potential use as a formative feedback tool to inform students' development of teamwork skills. A prominent feature of CATME and other peer evaluation tools is using quantitative measures that assess different dimensions of teamwork behaviors. These ratings can be reviewed by instructors and released to students so that they can gauge their performance across these dimensions of teamwork. The limitations of these ratings are that they are subject to various biases, such as friendships and personality perceptions, and impact on grades that can compromise the assessment's representativeness of actual individual performance [6].

Qualitative self and peer comments are a supplemental feature in CATME and other peer evaluation systems; they can provide valuable insights into teamwork behaviors by substantiating and explaining the quantitative ratings. While qualitative self and peer comments are a feature in these peer evaluation systems, instructors may refrain from releasing these peer comments to students because of potential harm and biases. Further, students may be apprehensive about providing critical feedback in qualitative peer comments during the semester if they know that instructors will release these comments to their peers, as it can potentially create a socially uncomfortable environment and negatively impact team dynamics [7]. However, withholding these peer comments to students may come at the

* Accepted 25 June 2024.

expense of providing valuable formative feedback that can support their individual professional development, especially in developing teamwork skills. Thus, there are nontrivial limitations with current peer evaluation systems that inhibit the extent to which students may receive formative feedback that can benefit teamwork skill development.

This paper explores harnessing generative artificial intelligence (AI) large language models (LLMs) to augment the peer evaluation process by mitigating the limitations of current peer evaluation systems' ability to provide students with high-quality, qualitative formative feedback. Specifically, we explore using AI-summarized Performance Feedback Reports (AI-PFRs) that use self and peer comments collected from CATME. The use of AI-PFRs in peer evaluation has several notable advantages. First, they provide greater anonymity to student peer comments, as individual comments are synthesized into AI-PFRs, which can improve student comfort in giving honest, critical feedback. Second, LLMs can potentially provide enhancements that improve the overall quality and organization of the original self and peer comments. This includes addressing harmful, unprofessional comments and comparing self and peer comments. These features can alleviate instructors' concerns about releasing unfiltered peer comments directly to students.

Overall, AI-PFRs are an opportunity to innovate the current peer evaluation process and provide students with timely and high-quality formative feedback that can support their development of teamwork skills. Our prior work [8] began to explore the potential for AI to effectively summarize self and peer comments into AI-PFRs, filter unprofessional comments, compare discrepancies between self-assessment and peer comments, and elaborate on peer comments to provide actionable feedback on how an individual can improve their teamwork behaviors.

This paper builds on our preliminary work by analyzing AI-PFRs created and implemented at scale in six first-year engineering PBL courses in the fall of 2023. We specifically explore the effectiveness of LLMs in improving the quality of the original peer comments through the following overarching research question: *To what extent does GPT-4 enhance the quality of feedback provided by student self and peer comments?*

Additional sub-questions that provide further granularity on these potential enhancements include:

1. To what extent does GPT-4 improve the quality of feedback to be more constructive and actionable compared to original peer comments?
2. How effective is GPT-4 in addressing inappropriate comments in original student peer feedback comments?
3. How effective is GPT-4 in comparing students' self-evaluation comments with peer comments in regard to their team performance?

Through these research questions, we can better understand the potential use of LLMs to enhance and summarize original peer comments with the purpose of supporting students' development of teamwork skills.

2. Literature Review

2.1 Theoretical Underpinnings

Feedback is an integral part of the learning process, as it allows students to reflect, identify areas for improvement, and increase motivation. While the context of peer feedback may differ, the benefits remain the same. Peer feedback can provide unique insights that may otherwise be overlooked. Bandura's Social Cognitive Theory of Self-Regulation posits that individuals learn and regulate their behavior through an interplay between personal factors, environmental influences, and cognitive processes. Self-regulation, a central concept of this theory, involves goal-setting, monitoring progress, and behavior adjustment to achieve desired outcomes [9]. Students can develop their self-regulated learning skills through formative assessments and feedback [10]. Regular access to feedback can stimulate self-regulation, with more frequent attention to feedback-standard gaps that encourage students to adjust their behavior and become more self-aware [11]. Considering the complexity of teamwork behaviors, multiple iterations of self-regulation and feedback are expected to lead to progressive improvements and mastery of these behaviors [12]. Thus, it is critical for student learning to receive multiple feedback assessments [13].

In PBL classes, instructors can play a significant role in organizing and implementing formative assessment and feedback opportunities to support self-regulated learning related to teamwork behaviors and skills [14]. However, in large PBL classes involving many teams, instructors are not immersed in the everyday team interactions and, consequently, are not well-positioned to individually assess and provide feedback on teamwork behavior to individual students. If students do not receive formative feedback on their teamwork behaviors, it can limit their ability to develop teamwork skills [15]. In our study, we focus on qualitative peer feedback, where students provide written feedback about each student's performance on the team. Peer feedback positively benefits students, instructors, and student teams in many ways and can help improve student learning [16]. Peer feedback provides an opportunity for students to receive formative and summative feedback. For example, formative feedback is provided throughout the semester and allows students to make improvements, while summative feedback is only provided at the end of the semester. Formative feedback has been identified as a way to help improve learning among students [17]. This is especially salient considering the importance of formative feedback in promoting the development of self-regulated learners [14].

In order to attain these benefits of feedback, studies have highlighted the importance of quality feedback. Principles of quality feedback have been highlighted by several researchers [12, 14, 18]. Shute [18] conducted a literature review on formative feedback in education and suggested that in order to improve learning, formative feedback should focus on the task, be specific, be as simple as possible, and orient to a learning goal. Similarly, Hattie and Timperly's model of feedback identifies four levels of feedback: task level, process level, self-regulation level, and self level [12]. Task level focuses on the performance of the tasks being completed, while the process level highlights what is needed to successfully perform the tasks at hand. The self-regulation level consists of students monitoring and regulating their actions, while the self level discusses personal evaluations of the learner. We developed the prompt for the AI-PFRs by considering the literature on self-regulated learning and effective feedback, which we further discuss in the Methods section.

In this literature review, we further elaborate on the benefits and limitations of peer feedback in classrooms, prior work on peer feedback in engineering education contexts, and the use of generative AI in engineering education contexts, including ethical considerations for its use in peer evaluation.

2.2 Peer Feedback in Student Teams

2.2.1 Benefits of Peer Feedback

Research has shown several benefits of peer evaluation systems in promoting student teamwork skills. For example, Brutus and Donia [13] found that using a peer evaluation system significantly improved the effectiveness of students working on teams. O'Neill, Boyce, and McLarnon [19] found that peer evaluations contribute to better team health and improved project grades. Moreover, Brutus, Donia, and Ronen [20] showed that using peer evaluations increases students' confidence in evaluating and providing feedback to their peers. Peer evaluation systems increase accountability and reduce social loafing among team members [3–5, 21], thus helping mitigate the likelihood that students are part of a dysfunctional team, which can undermine their potential to benefit from group work. Indeed, social loafing was the most frequently cited negative student team behavior in a review of engineering education articles and a source of conflict in student group work [22]. Harkins and Jackson [2] argue that the possibility of being evaluated by peers can motivate individuals to contribute more actively, reducing instances of social loafing. Similarly, formative peer feedback can help improve student motivation on student teams [20]. Research on an introductory design course found that students felt more motivated and accountable when they knew their peers would evaluate them. These findings collectively suggest the importance of peer evaluations as an accountability structure that mitigates social loafing, a major source of conflict in group work.

2.2.2 Limitations of Peer Feedback

Peer evaluation systems play a critical role in PBL contexts, but their full potential in providing students with high-quality, written feedback on their teamwork behaviors has yet to be fully realized. Current peer evaluation systems (e.g., CATME, ITP Metrics, TEAMMATES) contain quantitative rating scales that assess dimensions of teamwork behaviors with descriptive scales that are self-explanatory. However, they provide minimal to no instruction to students on how to write qualitative feedback, which likely negatively impacts the depth and specificity of the written feedback students receive. This lack of instruction and experience can lead to considerable variation in the quality of written feedback provided, as students may not be familiar with effective practices for writing feedback. Prior work has found that students cut and paste the same generic feedback for each of their peers, and many students do not typically provide enough specific detail in their feedback to be of value to the receiver [7].

Poor feedback can even lead to negative consequences, as evidenced by Kluger and DeNisi's [23] meta-analysis of feedback interventions. They found that feedback directed at the personal or ego level (e.g., feedback that praises or discourages individuals) can negatively affect performance. From our experience, even if students focus on task or process level feedback, they may be reluctant to identify areas for improvement or suggest ways in which improvement might be made. Students may also be hesitant to provide negative or critical feedback to their peers due to anticipated social discomfort, a prior experience that led to a negative impact on team dynamics, or fear of impacting a peer's academic record. Students who provide critical or negative feedback may not write it in an appropriate or constructive manner. For these reasons, instructors may summarize peer comments or screen them before releasing them to students, but this is especially challenging at scale. Alternatively, some instructors opt not to release peer comments at all, which prevents students from receiving formative feedback that can help inform their development of teamwork skills.

2.2.3 Peer Feedback in Engineering Education

A recent literature review about teamwork in engineering education identified feedback as a necessary component of building effective teams [24]. Although peer feedback is a critical component for teams in PBL classes, it can be a challenging process to facilitate because not all students are prepared to write effective feedback due to a lack of experience, training, or limited incentivization to do so. Recent work in engineering education has highlighted ways that peer feedback quality can be improved. Strategies such as instructional interventions [27] and collaborative peer review processes [28] have been shown to improve feedback quality, specificity, and breadth. Findings from a concurrent study on the impact of the feedback intervention used in this work showed a statistically significant increase in comment length and that feedback was found to be more actionable when students were trained on the elements of quality feedback [25]. When students were presented with examples of actionable peer feedback and able to practice, they were more likely to add detail, include both positive and negative feedback, and accompany identified gaps with recommendations for corrective action. These findings collectively demonstrate that training students on how to write feedback can improve the quality of feedback; however, the limitations described in the prior section remain.

2.3 Generative AI

Over the last year, AI language models have become a part of many people's daily routines. The use of LLMs has prompted increased discussion among researchers about generative AI's potential applications and its ethical utilization. Generally, recent papers in education underscore the potential advantages for educators using generative AI tools and potential drawbacks and limitations [26, 27]. In addition, many subfields of education have described potential uses of generative AI in teaching, learning, and research, such as chemical education [28], journalism and media education [29], and medical education [30].

In engineering education, a few papers have highlighted the uses of generative AI for engineering educators. A recent editorial in the Journal of Engineering Education described the potential uses of generative AI for research and teaching within the field [31]. Potential uses for research include generative assistance, data analysis, and research writing. Potential uses for teaching include assessment and teaching support. The authors highlight the importance of educators being trained to use this type of technology and the need for the community to work together to address potential concerns. Similarly, a recent conference paper used a metacognitive approach to have a conversation with ChatGPT on its potential uses in engineering education [32]. ChatGPT highlighted that instructors can use generative AI to create customized feedback and increase efficiency by automating certain tasks. Although there has been an increase in the number of people talking about generative AI in education, at the time this paper was written, few studies have showcased the use of generative AI in the classroom. Recent educational studies are increasingly exploring the use of generative AI tools in classroom settings. These studies involve applications such as aiding in physics lessons [33], assisting in problem development for prospective teachers [34], and evaluating the tool's impact on academic integrity and assessment standards [35].

Our study used GPT-4 to create personalized AI-PFRs for students working on a team-based semester-long project using team members' peer comments and their self-assessments. To use GPT-4 effectively in this context, we needed to mediate any potential harms of generative AI. In general, the limitations of AI have been the same across many fields; the most common include bias, lack of context, lack of common sense, and privacy concerns [27, 36]. A recent paper described that GPT-4 is limited because it lacks emotional intelligence [26], which is a significant consideration for our work. With our understanding and experiences with student feedback, there is the potential for peer comments to be unnecessarily rude or crass. If GPT-4 does not have the emotional intelligence to consider which comments may be hurtful, its potential use to summarize peer comments is limited. Su and Yang's paper provides a step-by-step framework for educators to appropriately use generative AI. They stress the importance of educators identifying desired outcomes, determining the appropriate level of automation, ensuring ethical considerations, and evaluating effectiveness [26]. This framework reflects our use of generative AI in engineering education classrooms, which we show in the methods section.

3. Methods

In our methods section, we share details on our participants, research site, data collection, data analysis, positionality, and study limitations. Process reliability and validity evidence are integrated throughout these subsections to illuminate potential influences on the study [37, 38]. Ethical considerations related to the use of AI, specifically LLMs, are also considered and discussed throughout these subsections.

3.1 Participants and Research Site

This study was conducted at a large R1 university in the mid-Atlantic region, where the College of Engineering makes up about 30% of the undergraduate student population. The participants in the study were all students in the first-year engineering program participating in a PBL class. In the PBL class, students work in teams on design projects for the entire semester. Students provide peer ratings and feedback to their teammates at multiple time points throughout the semester. Six classes participated in the pilot project involving three different instructors, and 303 students consented to their information being included in the study. In our study sample of 118, the gender distribution was reported as follows: 23% female, 73% male, and 2.5% either selected 'other' or preferred not to specify, with the remaining 1.5% not responding. Regarding race/ethnicity, 39.8% identified as White, 35.6% as Asian, 8.5% as Hispanic, 5.9% as Black, 1% as Middle Eastern or North African, 5.9% selected "Other," and 3.4% either declined to answer or did not respond.

3.2 Data Collection

Prior to the commencement of the study, we obtained approval from the university's Institutional Review Board (IRB) to ensure ethical compliance in our research practices. A training session about writing peer feedback was facilitated by one of the study authors and had several key aims: (1) emphasize the importance of peer feedback to learning and students' critical role in this process, (2) provide students with information on how to write effective peer feedback, and (3) introduce students to the research study. We recognize that the quality and effectiveness of the AI-PFRs are directly contingent upon the quality of the student's comments, underscoring the importance of this preparatory training. The session was designed to motivate students to write quality feedback (both positive and critical feedback) that focuses on specific teamwork behaviors. A particular emphasis was placed on encouraging students to provide specific examples of observed teamwork behaviors to support their comments and provide actionable feedback for critical comments. Details of this training and its impact on the peer comments can be found in our supporting work [25].

We collected the peer comments from each class after each training session at peer feedback time points, which were weeks 5, 7, or 8 of the semester, depending on the instructor. In each class, students were asked to write a self-evaluation and provide substantive comments for each of their team members in the CATME system. Prior to producing the AI-PFRs, we exported the data from CATME into an Excel sheet, compiled all of the peer comments, and then used a script to remove all names and pronouns from the comments. This allowed us to de-identify students from their data before sharing it with GPT-4 and also mitigated the introduction of any gender or other bias from GPT-4 based on assumptions related to the student's name or pronouns. The script added the prompt and combined all self and peer comments written for each student into an Excel cell. For each student, their prompt and de-identified self and peer comments were copied into GPT-4, which synthesized the comments into an AI-PFR.

We began the prompt by defining the context of the class and the generative AI model's role in providing formative feedback to students. The prompt then asks GPT-4 to summarize the comments and provide constructive, actionable feedback. Constructive feedback highlights areas for improvement and perceived weaknesses, and actionable feedback offers tangible suggestions for improvements. The prompt aligns with task and process-level feedback (refer to definitions in the Theoretical Underpinnings section), which are effective forms of formative feedback that can advance learning. The second part of the prompt asked GPT-4 to compare students' self-assessments to the peer comments. Considering students' self-assessments allows them to self-regulate their learning to improve their teamwork behaviors. Finally, the prompt concludes by asking to exclude inappropriate comments and output the report in letter format. There was one team that had less than two comments provided for some team members; since some of the team members did not complete the peer evaluations in time, we added a sentence that only generic feedback be provided to prevent feedback from being easily identifiable in these cases.

The following prompt was used to produce each AI-PFR:

> Prompt: "I am an engineering instructor who teaches a first-year engineering class where students work on teams for the semester. You are my assistant, who is going to help me provide formative feedback to my students. I collect peer comments periodically through-

out the semester, and I would like you to summarize the comments into a performance feedback review in a way that is constructive and actionable. Additionally, the students assess themselves and I would like you to compare their responses to the peer feedback. The output should be in the form of a letter, and please exclude anything that is inappropriate for the workplace." [If there are less than 2 comments for a student, please provide generic feedback only.]

This outlined process was used to produce an AI-PFR for all students across the six class sections. We copied each AI-PFR from GPT-4 to a Word document and re-inserted their names to the beginning of the letter (i.e., Dear [insert name]). Each AI-PFR was reviewed by the research team, and any edits were documented before it was shared with the instructor and student. We reviewed each report to ensure the content reflected the original peer comments and that there was no inappropriate content. We did not have students review their teammates' reports to ensure their raw comments were appropriately summarized and integrated. While there may be some benefits to doing this, we did not believe it was appropriate to have students also see other peer-to-peer comments that were not directed to them. We completed the process for each class over a week to provide timely feedback for all students. Data in this paper only includes comments and AI-PFRs from students who consented to their data being a part of the study. Further, we focused our analysis on student teams where every team member consented to the study, which resulted in 118 participants. We de-identified the data by replacing student names with pseudonyms. While we opted to utilize OpenAI's GPT-4 for this project, we acknowledge the existence of numerous alternative tools, including open-source options, that can be used for this purpose with the same or a similar prompt. We chose GPT-4 based on success in initial testing and found it was more effective for this task, especially for moderating inappropriate comments, in comparison to ChatGPT 3.5. We also note that OpenAI has recently partnered with a university, positioning it as a tool for educators [39].

3.3 Data Analysis

Our study employed a thematic analysis approach to identify patterns within the data. This method enabled us to interpret and categorize the data contained within the AI-PFRs and original student comments systematically and meaningfully. The thematic analysis process outlined by Braun and Clarke [40] informed our process. First, the authors read through all AI-PFRs to familiarize themselves with the data. Next, we conducted an open coding cycle on 118 AI-PFRs. This process involved detailed memo writing to capture insights about the general structure of the AI-PFRs and the primary differences between the original comments and the AI-PFRs, including noted improvements. Notable observations emerged at this stage on the structure of the AI-PFRs, which typically included a brief introduction, an outline of strengths and areas for improvement, a comparison of self and peer comments, and a final summary. We observed that the strengths and areas for improvement sections of the AI-PFRs effectively summarized the original comments, maintaining the core content but organizing it into coherent themes. This thematic sorting by GPT-4 enhanced the clarity and flow of the feedback while consolidating any redundant feedback. Moreover, the AI-PFRs often included actionable feedback, which was either minimally discussed or absent in the original comments. As a group, we discussed the emergent themes, which described the key characteristics of the AI-PFRs in comparison to the original comments.

Following the discussion and comparison of codes and memos, we decided to employ deductive coding based on the features of the prompt. The emergent themes we focused on include: (1) providing a summary of comments in a constructive, actionable manner, (2) addressing inappropriate comments, and (3) comparing self and peer feedback. A codebook was created with six distinct codes under our three thematic categories, aligning with our research sub-questions and the goals of the study. All the authors reviewed and provided feedback on the codebook before the second cycle of coding.

In the results section, representative examples from the coded AI-PFRs are discussed to provide evidence for each of the primary themes that align with our research questions. This approach not only underscores the effectiveness of GPT-4 in enhancing the quality of peer feedback but also illustrates how LLMs can address inappropriate comments and draw meaningful comparisons between self-evaluation and peer comments.

3.4 Positionality

The authors bring diverse experiences working in or managing engineering design teams in industry and academic contexts. Acknowledging how our backgrounds and experiences shape this qualitative research is essential to ensure an accurate representation of the data [41]. Through our previous experiences, we recognize the importance of collaboration in engineering work and the necessity to help students authentically practice and develop teamwork skills. In prior or current roles as engineers, engineering students, instructors, and teaching assistants (TAs), we have observed the limitations of the peer evaluation process from multiple perspectives and were aware of the challenge of providing students with actionable feedback on their teamwork behaviors. These experiences motivated our research study to pioneer a new approach using AI language models that can improve the peer feedback process and the quality of formative feedback students receive to support their professional development.

Similar to most educators in 2023, we were new to LLMs at the beginning of this project. We sought additional training in LLM tool prompting from our university to bolster the knowledge gained from current literature and our pilot study. While wary of potential risks to generative AI, we believe that tools like GPT-4 hold incredible potential for educators. To minimize the risks involved in using LLMs, we were intentional about protecting students' identities and mitigating potential gender biases by removing names and pronouns from the student comments. This practice was motivated by several of the author's experiences as women in engineering who have received biased and problematic feedback. The accuracy and quality of feedback tend to be even lower for women [42], which led us to seek innovative ways to provide screened, synthesized, and corroborated feedback reports.

4. Findings

In this section, we present a thematic analysis of the AI-PFRs. We used GPT-4 to process peer comments and generate comprehensive feedback reports in a letter format, as illustrated in Fig. 1. Each report encompasses multiple sections, including an introduction, an assessment of strengths, a discussion of areas for improvement, and concluding remarks. The LLM enhances the quality and flow of original comments, generating clear, reader-friendly reports. It processes inputs effectively, producing comprehensive feedback in letter format and organizing comments into coherent themes. It is important to note that the structure and content of these reports are adaptable and can vary based on the specific prompt used. Fig. 1 in this section illustrates how peer comments are transformed into the various components of the feedback report, with arrows indicating the flow from the initial comments to the final structured letter.

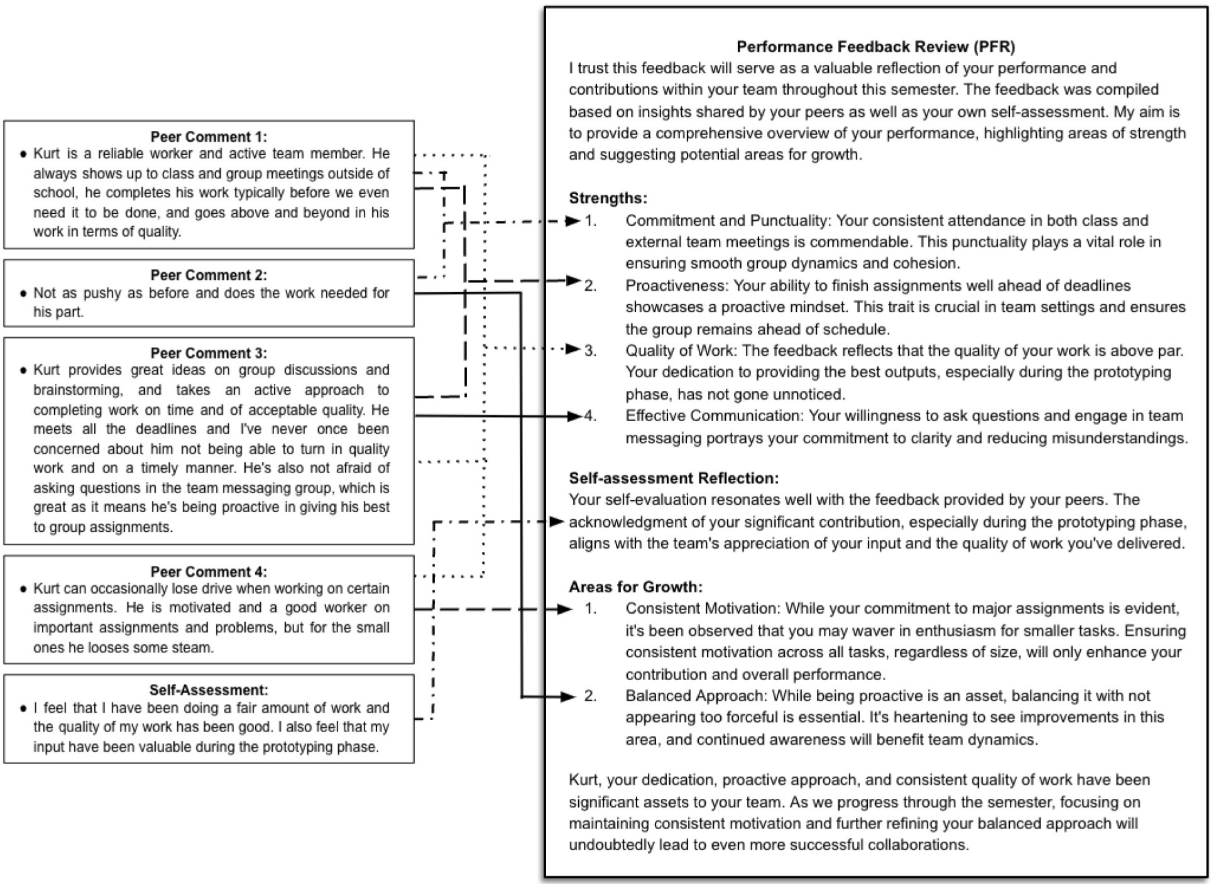

Fig. 1. The figure shows an example input (student comments) with the output (the PFR). We used arrows to illustrate how the comments mapped to the specific sections of the PFR. As you can see, GPT-4 analyzes the comments to provide students with feedback about their strengths, self-assessments, and areas for growth. Note: the comments were amended to remove names and pronouns when they were input to GPT-4, but for this illustration, we included a pseudonym and the original pronouns.

Every AI-PFR we generated was screened to ensure there was nothing harmful, unconstructive, or inappropriate. Throughout the screening process, we tracked any edits we made to the AI-PFRs and found our edits to be minimal. Most edits were formatting-related or related to the introduction, which could be changed by altering the prompt. We did a preliminary analysis of changes made to AI-PFRs, and we made edits to approximately 22% of them. However, we estimate that 85% of these changes were minimal and related to formatting, word choice, introduction preference, etc. The remaining changes were more substantial, which included rewriting a piece of feedback or rearranging the order of the feedback. We did not observe instances of feedback not mentioned in the original peer comments being introduced into the reports. Overall, very few substantial adjustments needed to be made to the AI-PFRs, which shows the reliability of using AI tools to create these types of documents.

The remaining Results section addresses how generative AI could enhance the quality of feedback provided by student and peer comments. Three primary themes (along with their underlying codes) address our three sub-research questions and are discussed further in the following subsections with supporting representative examples that replace actual student names with pseudonyms.

4.1 Improving the Quality of Constructive and Actionable Feedback

The AI-PFRs were able to both enhance constructive, actionable feedback from critical peer comments and enhance affirming feedback from positive peer comments. These two codes are discussed below.

4.1.1 Enhancing Constructive Feedback from Critical Comments

GPT-4 excelled in processing critical comments with minimal to no actionable feedback and adding specific, actionable feedback. Students often identified issues related to their peers' teamwork performance; however, they rarely suggested solutions, which AI-PFRs did effectively. This is highlighted in the following example in which Jayden's peers mentioned concerns about active and consistent participation in their peer comments (PC):

> PC: "Jayden is a mixed bag. On one hand he can be very helpful, on the other he could not be paying attention at all. It kind of just depends on the day."
>
> PC: "He's a moderate teammate with a strong knowledge base, however he lacks the necessary work ethic in my opinion. He sometimes doesn't contribute his thoughts to the group and kind of goes with the flow. However, I believe that in the future if he were to gain the drive, he could be a valuable team member."

The original feedback of being a "mixed bag" or a "moderate" teammate may not be well received, and although the feedback points out specific issues, it does not have clear next steps. The AI-PFR leverages these perhaps crass comments to provide actionable feedback for Jayden to consider, which was not present in the peer comment:

> "Enhanced Participation: While your relaxed demeanor is appreciated, ensure you strike a balance and actively share your ideas and thoughts during group discussions. Remember, your input is valuable.
>
> Seek Clarity Outside Class: You've mentioned sometimes feeling left out of certain discussions that occur outside of class hours. Take the initiative to approach teammates and request clarity or updates to ensure you're always in the loop.
>
> Consistency: While everyone has their off days, it's crucial to maintain a level of consistency in participation and contribution to ensure the team's success."

The AI-PFR still relays the critical feedback, but in a cordial manner with clear suggestions to the student that they should enhance participation and improve their consistency in participating. An additional example involves Cole, who did provide actionable feedback by writing the following, "I believe he could be more considerate of the feelings of others, since he sometimes is inconsiderate about laughing at someone explaining their ideas". The LLM was able to enhance this by making a connection to how this teamwork behavior can negatively impact communication through the following feedback:

> "Sensitivity: While humor can be a positive force, it's essential to be considerate of others' feelings. Laughing at someone's ideas, even unintentionally, can deter open communication. It would be beneficial to remain cognizant of this and to foster an environment where all ideas are respected."

Many of the examples showed trends toward tempering crass feedback, adding actionable feedback, and/or making connections to negative teamwork behaviors and how they impact the team.

4.1.2 Enhancing Affirming Feedback from Comments

The AI-PFRs were also effective in elaborating and enhancing the quality of positive peer comments. One illustrative example is Dylan, who only had two team members who completed a peer evaluation and wrote relatively brief comments. His teammates wrote the following peer comments):

> PC: "He is very encouraging for the team. He checks in to people to see if we are all doing well. Also, shows good effort for team project and interest."
>
> PC: "Dylan has an impressive depth of knowledge in engineering; he knows what he is doing. Additionally, he is a good communicator."

The AI-PFR organized and expanded these peer comments into a robust, detailed 'Strengths' section:

> "Supportive Nature: Your encouragement and regular check-ins with team members highlight your empathetic nature and commitment to team well-being. Such a supportive attitude fosters a healthy team environment, promoting collaboration and mutual respect.
>
> Proficiency in Engineering: Your peers have expressed their appreciation for your profound knowledge in engineering. Such expertise is invaluable in ensuring the project's success.
>
> Effective Communication: Effective communication is pivotal in teamwork, and you shine in this regard. This skill has a significant positive impact, ensuring everyone is on the same page and moving forward cohesively.
>
> Decision Making: Your ability to take the initiative and make decisions when required showcases your leadership qualities. Prompt and informed decision-making can be crucial, especially when deadlines loom."

This example illustrates how generative AI can produce a detailed feedback report based on limited information conveyed in the peer comments and connect positive attributes to how they impact the team. A detailed 'Strengths' section like this can be encouraging and confidence-boosting for students as it further elaborates and connects how a student's teamwork behaviors positively impact their team. Table 1 provides an additional example for each code for how GPT-4 enhances the quality of both critical and positive peer comments.

4.2 Addressing Irrelevant or Inappropriate Comments

Our prompt was designed to moderate inappropriate comments or potentially exclude them. Sometimes, peer feedback included comments that were entirely unrelated to teamwork behaviors or comments that were crass to the point that instructors may not be comfortable releasing them. In the following sections, we share examples of how the AI-PFRs addressed inappropriate or irrelevant comments.

4.2.1 Addressing Comments Unrelated to Teamwork Behaviors

While an opportunity for general feedback to the instructor was included in the CATME form, some students included course-related comments in their peer feedback. For example, Alexia noted how a teammate with the same name and similar minority identity was on the same team, which is not relevant to the team member on which they are commenting:

> "I'm guessing that Other [Alexia] is in my group because CATME's algorithm required itself to group minorities into pairs before forming groups. However, regardless of why they might be in my group, they are the most approachable team member in this entire group by a longshot (to me, at least)."

GPT-4 was able to ignore this mostly off-topic comment and focus on the relevant feedback, like the student's approachability. Another notable example of addressing an off-topic comment was when fantasy football was mentioned in a peer comment for Prem; the end of the AI-PFR included the note below:

> ["As an aside, while your proficiency in fantasy football might be a topic of light-hearted jest, remember that everyone has diverse strengths and expertise. Focus on what you bring to the academic table and continue to grow in those areas."]

The report identified that this was a joke within the team and was able to address it appropriately.

Table 1. Additional examples of how GPT-4 improved the quality of constructive and actionable feedback. The table provides both the original peer comments and the related AI-PFR excerpt

Code	Original Comments Excerpt	AI-PFR Excerpt
Enhancing Constructive Feedback from Critical Comments	**PC**: She has missed a few classes so sometimes she is lost on what the group needs to get done which makes her contributions a bit less overall. She also tends to join group meetings late, but is working on catching up. She always helps with the work, although she tends not to be around for zoom meetings. That is something that she could work on.	**Presence and Punctuality**: Your peers have mentioned that there have been instances where you've missed classes and joined group meetings late. We understand that unforeseen circumstances, such as illness, can affect participation. However, timely communication about your unavailability can aid in smoother team dynamics.
Enhancing Positive Feedback from Comments	**PC**: Donovan often is in charge of setting up meetings so we can get stuff done, and sending out zoom links. **PC**: He is very insightful and adds thorough ideas to the conversation which helps us build strong ideas for our team assignments. He is always excited and brings in the fun element with his sense of humor.	**Communication**: Your knack for communication stands out, and the team particularly appreciates your initiative in setting up meetings and ensuring a steady flow of information. **Insightfulness**: You bring depth to group discussions with your thorough ideas. This insight is essential in creating well-rounded and robust project outcomes. **Positive Atmosphere**: The balance you strike between seriousness and light-heartedness, with your sense of humor, is essential in creating a comfortable and enjoyable working environment.

4.2.2 Moderating Inappropriate Comments

Despite training on writing effective feedback, some students still wrote comments that instructors would not be comfortable directly releasing to students, or the comments used language that would not be constructive for or likely well-received by their peers. The LLM effectively removed or reworded these types of comments. This is important, as harsh language can be offensive and diminish any potential constructive elements of the feedback. For example, one of Alden's peers was expressing in their feedback that they sometimes voice their opinions over others and used the descriptor "dictator vibe":

> "Alden seems to have a mind of his own sometimes as he has the ideas and the passion for the engineering project, however, sometimes it overpowers our opinions. Voicing his opinion over ours has been common and it gives off a dictator vibe."

This feedback is not actionable nor constructive and could be perceived as offensive. It could negatively impact the working relationship if directly released to the student. When this was summarized with other feedback to create the AI-PFR using the prompt suggested, the following was mentioned on the topic within the report:

> "Leadership Dynamics: While your passion and diligence in the project are evident and admirable, there are concerns about occasionally overshadowing others' opinions. Balancing leadership with collaborative listening will enhance group synergy."

This part of the report still conveyed the issue raised by a peer but in a constructive way to the student. It also frames the concern in relation to the student's leadership abilities. Table 2 provides an additional example for each code for how the LLM addressed irrelevant or inappropriate comments.

4.3 Comparing Self Comments with Peer Comments

In comparing self comments with peer comments, the AI-PFRs would identify feedback discrepancies or affirm alignment in these comments.

4.3.1 Identifying Discrepancies between Self and Peer Comments

A powerful example of this discrepancy involves Sophie, who has taken an active leadership role on her team, but held reservations about how the other team members perceived her organizational skills. She explains in her self comment:

> "Oftentimes I am the one who checks on other group members and their progress in the current assignment in the group chat. I will also usually be the person who checks over all the work in the group document and I will send messages out to group members asking them to fix certain parts or elaborate a little more. I try to hold back on doing this though, because although I know that it's a group project and I'm just trying to work towards a good grade for the whole group I feel a little bossy when I do this."

Sophie's teammates did not see her as bossy, and instead gave very complimentary feedback about her work ethic. The AI-PFR addressed this discrepancy by reassuring Sophie that her skills are valued by the team:

> "Effective Communication: While you've noted your hesitations about appearing "bossy", your teammates value the clarity you provide, especially when checking over collective work or suggesting improvements. This shows you have the team's best interests at heart."

Furthermore, the AI-PFR provided actions that Sophie could take in the future if she felt unsure about group dynamics:

> "Seek Feedback Actively: Given your occasional hesitation in how you come across to teammates, consider actively seeking feedback after group tasks. This will help address any concerns and also provide insights into how best to approach collaborative work."

This example demonstrates how AI-PFRs can rebut internalized stereotypes found in self comments.

Table 2. Additional examples of how GPT-4 addressed irrelevant or inappropriate comments. The table provides both the original peer comments and the related AI-PFR excerpt. Self comments are shown as (SC) and peer comments as (PC)

Code	Original Comments Excerpt	AI-PFR Excerpt
Addressing Comments Unrelated to Teamwork Behaviors	**SC**: I decided coming into this class that I would not be committing a significant amount of time to it. I have taken this exact class before at [other college] and the credit did not transfer over for some reason I could not explain.	**Expanding Engagement**: While your contributions have been substantial, there's always room for further engagement. Going that 'extra mile', even in familiar territory, can yield unexpected learnings and deeper insights. **Share Past Experiences**: Since you've previously encountered this course content, your insights and experiences could be invaluable. Sharing more of these with the team might offer fresh perspectives and shortcuts.
Moderating Inappropriate Comments	**PC**: Savannah sometimes isn't fully mentally with us . . . Sometimes she does the bare minimum . . . which gets under some of our skin . . . Savannah is a good but inconsistent teammate.	**Focus & Engagement:** Feedback indicates moments of inattentiveness during group sessions. Ensuring consistent presence and active participation during team discussions can enhance collaboration and cohesion. **Task Quality:** While you consistently complete your assigned tasks, there's an expressed desire for a consistent higher level of effort, particularly in more complex assignments.

4.3.2 Affirming Alignment between Self and Peer Comments

An example of good alignment between self and peer comments involves Victor, whose team members recommended that he voice his ideas more:

> PC: "He is very quiet. I wish he would voice his opinion more, especially in group discussions as he tends to have good opinions and helpful perspectives. . ."
>
> PC: "Victor has a lot of great ideas, but generally prefers to work independently. Of course, he'll still speak out and share them verbally, just not as frequently as some other team members. Victor has a habit of working on other unrelated assignments during group discussions or assignments. . ."

In his self comment, Victor acknowledged a similar desire to engage further in group conversations:

> SC: ". . . However, I feel that it will be more beneficial in the long run if I were to be more engaged in the conversations that occur, rather than just being efficient at completing work. Although I do contribute one or two ideas to the conversations, I am rarely in control of the conversation as well."

The AI-PFR synthesized this consistent feedback by confirming the alignment between self and peer comments and encouraging Victor to engage in discussions:

> "Engagement in Discussions: Both your self-assessment and peer feedback indicate that there's room for improvement in your active participation during group discussions. Sharing your ideas and taking the lead in conversations can add further value to your team's collaborative efforts."

Table 3 provides an additional example for each code on how GPT-4 compared self and peer comments.

5. Discussion

In our study, we explored the potential of an LLM to enhance the quality of feedback in a first-year PBL context involving student teams. Our analysis concentrated on whether GPT-4 could produce feedback reports that enhance self and peer comments. More specifically, we sought to understand the extent to which GPT-4 could: (1) render feedback more actionable and constructive, (2) filter harmful or irrelevant feedback, and (3) create opportunity for self-awareness through comparing self and peer comments. These objectives were determined from literature in the education field

Table 3. Additional examples of how GPT-4 compared self and peer comments. The table provides both the original peer comments and the related AI-PFR excerpt

Code	Original Comment Excerpts	AI-PFR Excerpt
Identifying Discrepancies between Self and Peer Comments	**SC:** I feel I have performed well by contributing an equal amount of work to the group. I have made sure to speak up about any ideas or disagreements I have. Finally, I have made sure to communicate my schedule so we can meet at everyone's convenience. **PC1:** Does not talk too much unless prompted, and does not typically offer up his own ideas. Could be a little more open, as when he does speak, it is worth listening to **PC2:** Finn performs his work well and is often quite productive during team meetings, listening to conversation and filling out sections of whatever we're working on. However, Finn could do more to speak up and communicate his ideas with the group. He certainly has a unique perspective that deserves to be heard more often.	**Areas for Growth:** Active Participation: While your insights are valued when voiced, there's a shared feeling among your teammates that they would benefit from hearing more from you. Your perspective is unique and deserves to be shared more frequently. **Reflection on Self-Assessment:** Your self-perception is mostly aligned with the feedback from your peers. While you have identified your contributions and communication as strengths, there's an invitation from your team for more vocal participation in group discussions.
Affirming Alignment between Self and Peer Comments	**PC:** Ethan is very vocal and is always leading the conversation towards new ideas. . ..He is always asking questions and making suggestions. **PC:** Ethan has been a major contributor to the team thus far. He never dismisses anyone's ideas and is quick to make sure everyone is able to voice their feelings regarding ideas. . . **PC:** Organizes and plans a lot of our team meetings to make sure everyone's working and up to date on the project. Is also very encouraging towards everyone in the group. **SC:** I have made an effort to keep the team on track, planning times that we can meet and setting roadmaps for the work that must be completed. I try to get every team members' opinion on any major project decisions to ensure that everyone is comfortable with choices that may affect the team or the project. I respect every member of the team and their expertise or limitations, making sure that we are able to focus on our own individual skillsets, while still encouraging feedback from all members of the team and prompting conversations about our approach to problems and ways that our solutions could be improved.	Your self-assessment mirrors the sentiments of your peers. The effort you've made in keeping the team aligned, valuing every member's input, and acknowledging both strengths and areas of improvement within the team has been well-received. Your proactive approach in seeking feedback and initiating discussions to enhance solutions aligns perfectly with the feedback.

on the importance of quality feedback in promoting the self-regulation of behaviors. Our findings reveal that the AI-PFRs can effectively reach these objectives, resulting in immense potential for use in the PBL classroom.

Addressing our first sub-research question on feedback quality, we observed that AI-PFRs were adept at generating feedback linked to team behaviors and skills, even with sometimes limited input in the comments. This enhancement was particularly notable in making peer feedback comments more actionable, effectively refining both positive and critical comments to be more applicable and tied to how certain behaviors influence team dynamics. From our analysis, we observed many students gave peer feedback that was critical but not necessarily actionable. The AI-PFRs could draw from these critical comments to provide practical recommendations. Further, as shown in multiple examples, students would use terminology that could be seen as crass and may not be helpful to the reader, even if the sentiment were rooted in feedback that could be helpful. The ability to temper critical remarks and distill what would be useful was a strength of the AI-PFRs. This relates to our second sub-research question focused on addressing comments with inappropriate content, a valid concern for instructors in releasing peer feedback comments. We found that the AI-PFRs effectively identified and managed comments that may have been poorly received. Although instances of inappropriate comments were relatively rare, which could be partly attributed to prior student training given in class, we did not observe any instance when the AI-PFR included an inappropriate comment in a form that would be harmful or unconstructive to the student. Students also occasionally shared irrelevant comments, which were appropriately addressed.

Regarding our third sub-research question, the AI-PFRs synthesized self and peer comments, noting alignment and misalignment in peer feedback and self comments throughout the report or in a dedicated section. In most cases, students did have alignment between their self and peer comments. However, for some students, there were discrepancies between their self and peer comments, indicating some lack of self-awareness on how their teamwork behaviors are perceived. For students who may be over or under-confident about their teamwork skills, formative feedback that promotes awareness of teamwork behaviors can help them more effectively monitor and regulate their teamwork behaviors, which aligns with Bandura's Social Cognitive Theory of Self-Regulation [9].

This study builds upon existing literature and efforts in engineering education that solely focus on training students to provide better feedback [43, 44]. The quality of student self and peer comments is still critical to the quality of the AI-PFRs. However, training students to write better feedback does not fully address some of the challenges in improving the peer evaluation process discussed in the introduction and literature review sections. Moreover, the process of receiving the reports may help students see examples of how they can write feedback more effectively to improve their feedback writing skills. We designed our prompt to create a report that aligns with performance feedback reviews engineers may receive in the workplace as a way to promote useful and relevant learning for students who are less experienced in receiving peer feedback. Our findings suggest that GPT-4 augments the peer evaluation process by enhancing feedback quality, addressing irrelevant or inappropriate comments, and comparing self and peer comments, positioning it as a valuable pedagogical tool. We emphasize that the use of these reports complements, not replaces, the need for students to learn how to give and receive peer feedback. In the first-year PBL context, we see many strengths in the use of AI-PFRs; however, the goal is for this to be a stepping stone for students learning to openly share feedback with their peers constructively.

Overall, this study advocates for the potential of AI language models in overcoming the limitations of the peer evaluation process in PBL classes and improving the quality of feedback students receive. We recognize the benefits and potential challenges of receiving AI-summarized feedback. The use of AI-PFRs allows students to give and receive feedback without fear of teammates finding out exactly what they said, which may allow them to be more honest or transparent in their peer feedback comments. While we suspect this may be true and a valuable feature of this use of AI tools in the peer feedback process, students' perceptions of this were not explored in this study. It's possible some students may feel apprehensive about receiving personalized feedback from a non-human entity, which may lose the original tone or voice of the original peer comments. We also recognize that there are benefits to students openly discussing and managing conflicts directly with one another. While the AI-PFRs can potentially help inform and even catalyze these conversations, there may be drawbacks in students becoming overly reliant on using AI-PFRs as a means to safely communicate feedback and avoid having meaningful dialogue to resolve team issues. While pioneering in its application of AI tools in peer feedback for PBL classes, our research has limitations, which we will discuss in the subsection below.

5.1 Limitations

One limitation of our study stems from the training students received in providing peer feedback. As discussed, this influenced the quality and length of the peer comments compared to the prior iteration of the course [25], which directly influences the quality of the AI-PFRs. We strongly urge instructors who are interested in implementing AI-PFRs in their courses to provide guidance to students on how to write self-reflection and peer feedback comments. Assignment descriptions with specific instructors and a rubric to assess the quality of self and peer comments can help ensure students provide enough content to inform the AI-PFRs.

Finally, our study's deductive approach to analysis focused on the tool's effectiveness in producing quality feedback based on established criteria from the literature. However, preliminary analyses indicate potential gender differences in feedback, suggesting an avenue for future research. We also acknowledge that the framing of our prompts, as detailed in the methods section, is influenced by researcher decisions, which may guide subsequent iterations of this work. Our findings were meant to show proof of concept of the use of LLMs in PBL classes related to peer evaluations; therefore, our findings were scoped to align with the research questions.

5.2 Future Work

This study introduces an innovative tool to aid instructors in providing timely and actionable feedback. To build upon the positive initial results presented here, we propose further exploration into the use and impact of AI-PFRs. Building upon RQ2, analysis of how inappropriate feedback differs by gender and how these reports can reduce bias is one area of interest. Further, future work could explore how students perceive the AI-PFRs and their integration into the classroom, such as investigating optimal timing of feedback, frequency of feedback, and methods for integrating these reports into the learning environment. This may include how to integrate complementary pedagogical approaches, such as individual reflections and team debriefs, with the AI-PFRs that encourage students and teams to meaningfully reflect on their teamwork and involve understanding how the reports influence students' regulation of teamwork behaviors and their development of teamwork skills, including their ability to write and receive feedback. As students have more practice and exposure to the peer feedback process, they may become more comfortable with and even prefer approaches to providing more direct feedback, whether that is directly releasing peer evaluation comments or having team conversations. Exploring how instructors can effectively scaffold how students provide and receive feedback in PBL classes to promote teamwork skill development is an area of future exploration. Finally, we did not alter any of GPT-4's settings, which is an area for future work. Notably, the variety of wordings observed at the beginning of the AI-PFRs may have been minimized by adjusting the "temperature" parameter, which controls the model's randomness. Moreover, our findings are based solely on the use of GPT-4, and further research is needed to determine if these results are consistent across other LLMs.

6. Conclusion

Our study presents AI-PFRs as a novel educational innovation that enhances traditional peer feedback processes. We offer our findings as a foundation for other problem-based learning (PBL) environments looking to integrate such practices, recognizing the dynamic and evolving nature of this digital frontier. Through our analysis, we have established that LLMs can effectively summarize self and peer feedback for student teams, elevating existing educational tools and reaching new digital frontiers in how students engage in PBL courses. Addressing our research questions, we conclude that LLMs significantly enhance the quality of peer feedback when it comes to making feedback more constructive and actionable, addressing off-topic and inappropriate comments, and meaningfully comparing students' self-evaluation with their peer comments. We offer this work as a foundation for considering the use of generative AI in the PBL classroom, with the hope of giving students more opportunities for feedback so that they can refine their teamwork skills.

References

1. H. J. Passow and C. H. Passow, What Competencies Should Undergraduate Engineering Programs Emphasize? A Systematic Review, *J. Eng. Educ.*, **106**(3), pp. 475–526, 2017.
2. S. G. Harkins and J. M. Jackson, The Role of Evaluation in Eliminating Social Loafing, *Pers. Soc. Psychol. Bull.*, **11**(4), pp. 457–465, 1985.
3. C. M. Brooks and J. L. Ammons, Free Riding in Group Projects and the Effects of Timing, Frequency, and Specificity of Criteria in Peer Assessments, *J. Educ. Bus.*, **78**(5), pp. 268–272, 2003.
4. K. H. Price, D. A. Harrison and J. H. Gavin, Withholding inputs in team contexts: Member composition, interaction processes, evaluation structure, and social loafing, *J. Appl. Psychol.*, **91**(6), pp. 1375–1384, 2006.

5. D. Hall and S. Buzwell, The problem of free-riding in group projects: Looking beyond social loafing as reason for non-contribution, *Act. Learn. High. Educ.*, **14**(1), pp. 37–49, 2013.
6. C. O. Mayfield and J. R. Tombaugh, Why peer evaluations in student teams don't tell us what we think they do, *J. Educ. Bus.*, **94**(2), pp. 125–138, 2019.
7. A. Burgess, C. Roberts, A. S. Lane, I. Haq, T. Clark, E. Kalman, N. Pappalardo and J. Bleasel, Peer review in team-based learning: influencing feedback literacy, *BMC Med. Educ.*, **21**(1), p. 426, 2021.
8. S. Sajadi, O. Ryan, L. Schibelius and M. Huerta, WIP: Using Generative AI to Assist in Individual Performance Feedback for Engineering Student Teams, in *2023 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, 2023.
9. A. Bandura, Social cognitive theory of self-regulation, *Organ. Behav. Hum. Decis. Process.*, **50**(2), pp. 248–287, 1991.
10. I. Clark, Formative Assessment: Assessment Is for Self-regulated Learning, *Educ. Psychol. Rev.*, **24**(2), pp. 205–249, 2012.
11. B. J. Zimmerman, Becoming a Self-Regulated Learner: An Overview, *Theory Pract.*, **41**(2), pp. 64–70, 2002.
12. J. Hattie and H. Timperley, The Power of Feedback, *Rev. Educ. Res.*, **77**(1), pp. 81–112, 2007.
13. S. Brutus and M. B. L. Donia, Improving the Effectiveness of Students in Groups With a Centralized Peer Evaluation System, *Acad. Manag. Learn. Educ.*, **9**(4), pp. 652–662, 2010.
14. D. J. Nicol and D. Macfarlane-Dick, Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Stud. High. Educ.*, **31**(2), pp. 199–218, 2006.
15. J. Qadir, A. E. M. Taha, K. L. A. Yau, J. Ponciano, S. Hussain, A. Al-Fuqaha and M. A. Imran, Leveraging the force of formative assessment & feedback for effective engineering education, 2020, Accessed: Sep. 28, 2023. [Online]. Available: https://eprints.gla.ac.uk/202106
16. L. Li, X. Liu and A. L. Steckelberg, Assessor or assessee: How student learning improves by giving and receiving peer feedback, *Br. J. Educ. Technol.*, **41**(3), pp. 525–536, 2010.
17. A. Irons and S. Elkington, *Enhancing Learning through Formative Assessment and Feedback*, Routledge, 2021.
18. V. J. Shute, Focus on Formative Feedback, *Rev. Educ. Res.*, **78**(1), pp. 153–189, 2008.
19. T. A. ONeill, M. Boyce and M. J. W. McLarnon, Team Health and Project Quality Are Improved When Peer Evaluation Scores Affect Grades on Team Projects, *Front. Educ.*, **5**, 2020, Accessed: Oct. 05, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/feduc.2020.00049
20. S. Brutus, M. B. L. Donia and S. Ronen, Can Business Students Learn to Evaluate Better? Evidence From Repeated Exposure to a Peer-Evaluation System, *Acad. Manag. Learn. Educ.*, **12**(1), pp. 18–31, 2013.
21. N. Mentzer, A. Jackson, K. A. Richards, A. N. Zissimopoulos and D. Laux, Student Perceptions on the Impact of Formative Peer Team Member Effectiveness Evaluation in an Introductory Design Course, presented at the *2015 ASEE Annual Conference & Exposition*, Jun. 2015, p. 26.1422.1-26.1422.20. Accessed: Sep. 06, 2023. [Online]. Available: https://peer.asee.org/student-perceptions-on-the-impact-of-formative-peer-team-member-effectiveness-evaluation-in-an-introductory-design-course
22. M. Borrego, J. Karlin, L. D. McNair and K. Beddoes, Team Effectiveness Theory from Industrial and Organizational Psychology Applied to Engineering Student Project Teams: A Research Review, *J. Eng. Educ.*, **102**(4), pp. 472–512, 2013.
23. A. N. Kluger and A. DeNisi, The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory, *Psychol. Bull.*, **119**(2), pp. 254–284, 1996.
24. T. Chowdhury and H. Murzi, Literature review: Exploring teamwork in engineering education, in *Proceedings of the Conference: Research in Engineering Education Symposium, Cape Town, South Africa*, 2019, pp. 10–12. Accessed: Sep. 28, 2023. [Online]. Available: https://www.researchgate.net/profile/Homero-Murzi/publication/334681127_Literature_Review_Exploring_Teamwork_in_Engineering_Education/links/5e3b133da6fdccd9658a79c4/Literature-Review-Exploring-Teamwork-in-Engineering-Education.pdf
25. K. Drinkwater, O. Ryan, S. Sajadi, M. Huerta and M. Fisher, Improving Peer Feedback in Project-Based Learning Contexts: An Investigation into a First-Year Engineering Intervention, in *2024 ASEE Annual Conference & Exposition*, Jun. 2024.
26. J. Su and W. Yang, Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education, *ECNU Rev. Educ.*, **6**(3), pp. 355–366, 2023.
27. D. Baidoo-Anu and L. O. Ansah, Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning, *J. AI*, **7**(1), Art. no. 1.
28. E. A. Alasadi and C. R. Baiz, Generative AI in Education and Research: Opportunities, Concerns, and Solutions, *J. Chem. Educ.*, **100**(8), pp. 2965–2971, 2023.
29. J. V. Pavlik, Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education, *Journal. Mass Commun. Educ.*, **78**(1), pp. 84–93, 2023.
30. T. Jowsey, J. Stokes-Parish, R. Singleton and M. Todorovic, Medical education empowered by generative artificial intelligence large language models, *Trends Mol. Med.*, **29**(12), 2023.
31. A. Johri, A. S. Katz, J. Qadir and A. Hingle, Generative artificial intelligence and engineering education, *J. Eng. Educ.*, **112**(3), pp. 572–577, 2023.
32. J. Qadir, Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education, in *2023 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1–9. 2023.
33. P. Bitzenbauer, ChatGPT in physics education: A pilot study on easy-to-implement activities, *Contemp. Educ. Technol.*, **15**(3), p. ep430, 2023.
34. S. Küchemann, S. Steinert, N. Revenga, M. Schweinberger, Y. Dinc, K. E. Avila and J. Kuhn, Physics task development of prospective physics teachers using ChatGPT, *Phys. Rev. Phys. Educ. Res.*, **19**(2), p. 020128, 2023.
35. S. Nikolic, S. Daniel, R. Haque, M. Belkina, G. M. Hassan, S. Grundy, S. Lyden, Neal, P. and Sandison, C., ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity, *Eur. J. Eng. Educ.*, **48**(4), pp. 559–614, 2023.
36. F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau and L. Chen, Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration, *J. Inf. Technol. Case Appl. Res.*, **25**(3), pp. 277–304, 2023.
37. N. Kellam and A. M. Cirell, Quality Considerations in Qualitative Inquiry: Expanding Our Understandings for the Broader Dissemination of Qualitative Research, *J. Eng. Educ.*, **107**(3), pp. 355–361, 2018.

38. J. Walther, N. W. Sochacka, L. C. Benson, A. E. Bumbaco, N. Kellam, A. L. Pawley and C. M. Phillips, Qualitative Research Quality: A Collaborative Inquiry Across Multiple Methodological Perspectives, *J. Eng. Educ.*, **106**(3), pp. 398–430, 2017.
39. T. Swaak, Arizona State and OpenAI Are Now Partners. What Does That Mean? *The Chronicle of Higher Education*, Jan. 19, 2024. Accessed: May 1, 2024. [Online]. Available: https://www-chronicle-com/article/arizona-state-and-openai-are-now-partners-what-does-that-mean.
40. V. Braun and V. Clarke, Using thematic analysis in psychology, *Qual. Res. Psychol.*, **3**(2), pp. 77–101, 2006.
41. S. Secules, C. McCall, J. A. Mejia, C. Beebe, A. S. Masters, M. L. Sánchez-Peña and M. Svyantek, Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community, *J. Eng. Educ.*, **110**(1), pp. 19–43, 2021.
42. E. Doldor, M. Wyatt and J. Silvester, Research: Men Get More Actionable Feedback Than Women, *Harvard Business Review*, Feb. 10, 2021. Accessed: Dec. 08, 2023. [Online]. Available: https://hbr.org/2021/02/research-men-get-more-actionable-feedback-than-women
43. W. Huang, R. Wynkoop, M. Exter and F. Berry, Feedback Matters: Self-and-Peer Assessment Made Better with Instructional Interventions, presented at the *2022 ASEE Annual Conference & Exposition*, Aug. 2022. Accessed: Sep. 06, 2023. [Online]. Available: https://peer.asee.org/feedback-matters-self-and-peer-assessment-made-better-with-instructional-interventions
44. M. Mandala, C. Schunn, S. Dow, M. Goldberg, J. Pearlman, W. Clark and I. Mena, Impact of Collaborative Team Peer Review on the Quality of Feedback in Engineering Design Projects, *International Journal of Engineering Education*, **34**(4), pp.1299–1313, 2018.

Susan Sajadi (she/her) is an Assistant Professor in the Department of Engineering Education at Virginia Tech. She has a BS and MS in Biomedical Engineering. She worked as a biomedical engineer and later in corporate social responsibility related to STEM outreach before pursuing a PhD in Engineering Education Systems and Design from Arizona State University. Her research focuses on advancing inclusion, diversity, and equity within engineering educational settings and the engineering workplace.

Mark Huerta (him/his) is an Assistant Professor in the Department of Engineering Education at Virginia Tech. He received his BS and MS in Biomedical Engineering and PhD in Engineering Education Systems and Design from Arizona State University. His research interests include exploring approaches to cultivating positive, inclusive learning environments and the development of professional skills, including teamwork, in engineering education contexts.

Olivia Ryan (she/her) is a PhD candidate in the Department of Engineering Education and a Master's student in Engineering Mechanics at Virginia Tech. She holds a BS in engineering from Roger Williams University. Prior to starting school at VT, she worked as an electrical engineer in the power industry and at an education non-profit working in Chicago Public Schools. Her research interests include characterizing curriculum barriers in engineering related to mathematics and helping engineering students develop professional skills.

Katie Drinkwater (she/her) is a first-year PhD student in Engineering Education at Virginia Tech. She holds a BS in Mechanical Engineering from Duke University. Her research interests include engineering design pedagogy, extra-curricular involvement, makerspaces, and women in engineering.