How Guiding Questions Facilitate Feedback Exchange in Project-Based Learning

Amy Cook
Carnegie Mellon University
Pittsburgh, Pennsylvania
amyshann@andrew.cmu.edu

Jessica Hammer
Carnegie Mellon University
Pittsburgh, Pennsylvania
hammerj@cs.cmu.edu

Salma Elsayed-Ali
William and Mary
Williamsburg, Virginia
shelsayedali@email.wm.edu

Steven Dow
University of California San Diego
San Diego, California
spdow@ucsd.edu

ABSTRACT

Peer feedback is essential for learning in project-based disciplines. However, students often need guidance when acting as either a feedback provider or a feedback receiver, both to gain from peer feedback and to criticize their peers' work. This paper explores how to more effectively scaffold this exchange such that peers more deeply engage in the feedback process. Within a game design course, we introduced different processes for feedback receivers to write questions to guide peer feedback. Feedback receivers wrote four main types of guiding questions: improve, share, brainstorm, critique. We found that "improve" questions tended to lead to better feedback (more specific, critical, and actionable) than other question types, but feedback receivers wrote improve questions least often. We offer insights on how best to scaffold the question-writing process to facilitate peer feedback exchange.

CCS CONCEPTS

• **Applied computing** → **Collaborative learning**; • **Human-centered computing** → *Computer supported cooperative work*.

KEYWORDS

Project-based learning, peer feedback systems.

1 INTRODUCTION

Peer feedback exchange is an essential activity for project-based learning, especially in open-ended and iteration-driven domains like design [3][9]. Providing peer feedback to others helps feedback providers learn to evaluate work and understand the instructor's standards [28]. Receiving feedback from others helps feedback receivers develop self-assessment abilities and learn to use feedback to inform their iterative design choices [6][19].

However, there are several known challenges to peer feedback exchange. First, the quality of peer feedback is highly variable. Some students give "better" feedback than others, and feedback receivers may not know how to distinguish good feedback from bad. Second, one goal of peer feedback exchange is to help students align with instructor standards, but this goal is not always met. In order for students to reach that goal, they need to engage in meta-cognitive and reflective processes to self-assess their own learning and performance [5][13]. However, prior work shows that students often struggle to reflect on feedback even when prompted [9][17].

One method instructors use to address these challenges is to provide rubrics for feedback providers to use when they give feedback to their classmates. These instructor-authored questions are meant to scaffold the feedback process and guide students towards providing high quality feedback that is specific, critical, and actionable [20][22][27]. However, even when instructor-authored rubrics are used, feedback providers sometimes still struggle to write specific, critical, and actionable feedback for their peers [17].

This paper investigates the effect of a new method for addressing these challenges: student-authored guiding questions. We implemented guiding questions in a college-level educational game design course. The feedback process was conducted using an existing digital peer feedback system. We identified and analyzed several indicators of effective and thoughtful feedback exchange. We developed a qualitative coding scheme for what kinds of questions feedback receivers write. We analyzed how much and what kind of feedback was elicited by these questions. We used self-report data to determine which questions and feedback were most valued by feedback receivers and instructors. We also investigated how other factors, such as the question-writing process (individual or collaborative), may have impacted the types of questions feedback receivers wrote.

We found that improve questions tended to produce more specific, critical, and actionable feedback than other questions types. Also, feedback receivers who wrote questions collaboratively learned to internalize the instructor's values and standards, and to recognize the value of improve questions, while feedback receivers who wrote questions individually valued other less effective question types instead. This work provides preliminary support for implementing student-authored guiding questions as an interactive learning activity to scaffold more effective peer feedback exchange. We discuss implications for in-class learning technology, key open questions, and future work.

2 RELATED WORK

Peer feedback plays a pivotal role in project-based learning and design education, providing students with the opportunity to learn from one another and to iterate on and improve their projects [19][20]. Receiving peer feedback is also a critical component of self-regulated learning [6], and giving peer feedback to others can facilitate learning that transfers to a student's own project [18]. Both giving and receiving peer feedback offers a unique learning opportunity for students to become better acquainted with self-assessment, communication, and project improvement [6][26].

Methods for Implementing Peer Feedback

While some classrooms opt for verbal peer feedback exchange, others elect to use a written format either by hand or digitally when exchanging peer feedback. These methods have trade offs.

First, verbal peer feedback enables all members of a group to hear the feedback, but can simultaneously prevent everyone in the group from participating and can easily be forgotten unless recorded [24]. Second, exchanging traditional paper-based peer feedback allows for tangibility and flexibility, but may harbor logistical challenges come time

for collection and distribution [15]. Third, digital peer feedback systems are a relatively novel way of exchanging peer feedback [9][23][25]. Digital feedback exchange can be as effective as handwritten feedback, encourage more peer-to-peer interaction, and offer logistical and organizational advantages. However, digital methods of exchanging peer feedback may be distracting and challenging to adopt [24]. Using technology during peer feedback exchange has not only enabled peer feedback, but also gives us a new lens for understanding the dynamics of feedback exchange.

Qualities of Successful Peer Feedback

Successful peer feedback is specific, critical, and actionable [20][22]. Prior researchers developed a set of conditions that detail how formative assessments contribute to and have the potential to improve student learning; they state that feedback has to be specific, detailed, and facilitative in order for it to be useful [13]. Successful feedback also indicates the shortcomings of a project or provides a critical judgment on the handling of a particular aspect of a project [16]. Care should be taken to be critical of students' performance and content rather than personal characteristics [13].

Feedback should also provide an opportunity for feedback receivers to close the gap between their current performance and the desired outcome [20], and should motivate feedback receivers to act upon the feedback received, thus improving their work and increasing their learning [13].

For this study, we use those criteria - specific, critical, and actionable - to evaluate peer feedback quality. Other factors also influence the success of peer feedback exchange, such as the quantity of feedback received [26] and the amount of time that passes before feedback is received [17].

Challenges to Successful Peer Feedback

Despite the substantial value that can be derived from peer feedback, several challenges still exist during peer feedback exchange. First, peer feedback quality depends on the expertise of the person providing feedback [8]. Experts frame and analyze problems differently than novices. Thus, students may not be equipped to provide specific peer feedback, and peer feedback quality can be unreliable.

Second, students may feel anxious or hesitant to provide feedback on peers' work, especially when being critical [10][16]. However, prior work shows that constructive criticism positively impacts feedback quality and should be encouraged.

Third, peers may choose not to accept feedback especially if it is critical, potentially viewing it as invalid or as an inaccurate representation of their work [26]. Peers may not view each other as "knowledge authorities" but rather as

equal-status learners, so they may be hesitant to accept judgment or advice from peers [14]. Therefore, students may not be open or receptive to criticism of their work.

Using Rubrics to Guide Peer Feedback Exchange

Instructors sometimes use rubrics to scaffold the peer feedback process in hopes of addressing the above challenges. A rubric is "a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor" [2]. Rubrics can help novices frame their feedback according to the evaluation criteria, and can encourage peers to provide critical feedback. Prior work has investigated how students can use rubrics not just to evaluate, but also to learn. In particular, students reported using rubrics to "guide or reflect on feedback from others" [2], and existing peer feedback systems have successfully used structured rubrics authored by the instructor to guide feedback [17][25][27].

Students may learn more from authoring their own rubrics than from using a general rubric authored by the instructor for the entire class [7]. Student-authored rubrics may help increase a student's sense of ownership over the feedback, giving them a stronger motivation to engage with critique. Student-authored rubrics also provide opportunities to both increase reflection about what type of feedback is desired and guide feedback providers. These opportunities drive our investigation into how rubrics can best be implemented with in-class peer feedback systems.

ICAP Framework for Types of Learning Activities

ICAP is a theoretical framework that suggests guidelines to improve student engagement in learning activities [7]. It classifies learning activities into four types: *interactive, constructive, active,* and *passive.* As students transition from being passively engaged to interactively engaged, their levels of learning should also increase.

Using the example of a lecture, students receiving and digesting information without overtly doing anything else would be considered a passive learning activity. To make the lecture into an active learning activity, students would need to manipulate the information they received. For example, they might take notes. A constructive learning activity, which promotes deeper learning gains than active or passive activities, would be where students generate some output of knowledge in response to the information received, such as formulating and asking a question or drawing a concept map. However, constructive activities are done individually, which restricts mutual exchange and the potential generation of new ideas. Thus, an interactive learning activity, where students engage in knowledge creation (a constructive learning activity) *while discussing with their peers,* would produce the highest learning gains according to ICAP theory.

ICAP theory might help explain why feedback providers struggle to use instructor-authored rubrics. Understood through the lens of ICAP theory, using instructor-authored questions to guide feedback providers is an active learning activity, because feedback receivers are engaging with questions they received from the instructor, but not constructing something new. To make this aspect of peer feedback process constructive, feedback receivers could produce the questions themselves instead of simply using instructor questions. To make the process interactive, feedback receivers could write these guiding questions collaboratively with their teammates instead of individually.

Writing guiding questions has the potential to help feedback receivers invest more and to think more critically about what they need from the feedback process [7]. However, we know little about how students write questions for peer feedback and how these questions affect the responses from peers and instructors.

This work investigates the process of engaging students in writing questions to guide the peers providing feedback. We used ICAP theory to inform the design of our interventions. Asking students to prepare guiding questions for peer feedback exchange raises the following research questions:

(1) What kinds of guiding questions do students write?
(2) (How) do students and instructors respond differently?
(3) How do the guiding questions affect the feedback?

To explore these questions, we conducted a study where teams in a project-based course wrote questions for their peers for a feedback-exchange activity, and we analyzed how students wrote and responded to questions.

3 METHOD

Participants

34 graduate students and 1 instructor in a project-based educational game design course at a US university participated in a 6-week-long study. Students were assigned to a team of approximately 3 students to complete an educational game design project. Students remained on the same team for the duration of the project. The 10 teams chose their own team name and their project topic.

Platform

We used a digital peer feedback system, called PeerPresents, to facilitate feedback exchange. This system was designed for open-ended in-class feedback on team presentations [23][24]. In this system, feedback receivers can write questions to guide feedback providers. The system generates a URL that can be shared with the feedback providers. Feedback providers can pick any username to use within the system. They provide feedback by typing into a text box to answer each of the presenters' questions. They can also
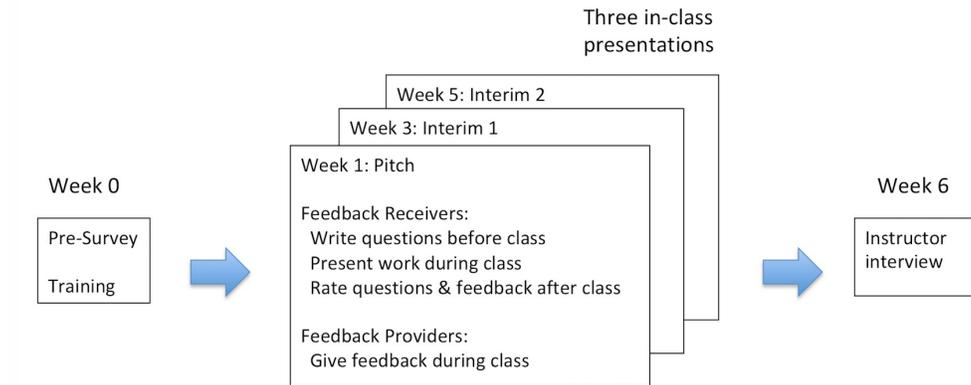
Figure 1: Students completed study tasks before and after each in-class feedback exchange.

read and react to feedback provided by others. Afterwards, the system's reflection interface allows teams to read and organize the feedback they have received. These last two features (reactions and the reflection interface) are novel to this peer feedback system.

Initial pilot evaluations revealed this system allows students to receive immediate and diverse feedback, and students reported providing more feedback with this system than other methods [23]. Further evaluation showed this system to be as effective as a paper-based feedback exchange in terms of feedback quality and quantity, but the digital system provided additional ways for students to participate and required less work from the instructor [24].

Procedure

Figure 1 outlines the study procedure and indicates when students completed each study task. As an introduction, all study participants completed a pre-survey about their prior experiences and attitudes towards giving and receiving peer feedback digitally. They were then trained during class on best practices of writing feedback. Participants were asked to identify what they felt were qualities of helpful peer feedback, then shown research that indicates specific, actionable, and critical feedback is most helpful. Participants were shown "unhelpful" feedback comments on a mock presentation and asked to revise the comments to make them more specific or more actionable. Participants were also shown examples of questions they might use to elicit feedback, and asked to turn various forms of "yes or no" questions into open-ended questions. A detailed description of the training is provided as supplementary material.

The teams presented demos of their game design projects in class every other week during the six-week project, for a total of three presentations: pitch, interim 1, and interim

2. On presentation days, each student would act as both a feedback provider and a feedback receiver.

As feedback receivers, students completed three tasks. First, the day prior to presenting, each team created a set of three guiding questions that would be displayed in the digital peer feedback tool to guide the feedback they would receive on their project. Teams could choose whether to write their guiding questions collaboratively as a team or individually. Four teams chose an individual process (which would be defined as a constructive activity in the ICAP framework) and six teams chose a collaborative process (interactive). We observed one team's collaborative question-writing session. One researcher sat with three team members while they wrote and discussed questions for 23 minutes. The researcher took notes on a laptop about what questions students wrote and why students revised or rejected certain questions.

Second, teams would present their game demo to their peers during class on each presentation day. Because all 10 teams presented each class period, teams were given only 5 minutes to present their work.

Third, after class ended, feedback receivers rated their team's guiding questions and the individual feedback comments received in response to their questions on a scale from 1 to 5. Feedback receivers did not rate questions written by other teams or feedback given to other teams. Ratings were collected using an online survey.

As feedback providers, students had only one task. During each in-class presentation, feedback providers used the digital peer feedback tool to give feedback to their peers, by answering the guiding questions written by the feedback receivers.

The instructor was asked to rate each team's questions and a subset of the feedback received by each team after each presentation. The instructor was also briefly interviewed

after the course ended about his experience with this peer feedback process.

Data Analysis

We analyzed student and instructor responses to the surveys and all the questions and comments provided via the digital peer feedback tool. We assessed the quality of guiding questions by coding for question type and analyzing the quality of feedback generated by each question type. We assessed feedback quality by coding for specificity, critical sentiment, and actionability, as prior research highlighted these factors as markers of effective feedback. We present descriptive statistics to portray what is going on in this data. We used standard error of the mean to measure variability. We do not present inferential statistics, as this exploratory study did not produce enough data for an accurate hierarchical model.

Analyzing the Type of Guiding Questions. We developed a coding scheme that categorized questions into 4 distinct types. To define these types, we first took a random sample of 50 student-authored guiding questions and used a card-sort method to categorize them. We then checked our codes against literature that classifies questions for teacher training purposes [4][12][21] to see if we had missed any major question categories. Three researchers practiced coding on an old data set using the categories derived from the first round of card sorting. The researchers reached agreement of 84% when coding for question type (see Table 1). To calculate percent agreement, we divided the number of questions where all three researchers agreed on the code for questions type by the total number of questions that were coded.

Question Type	Examples
Brainstorm	"Any ideas about how to go about scoring the game? EX: Individual winner, ranking players (1,2,3) ...."
Critique	"What do you think about the design of our game board and share cards?"
Improve	"How do we make this game less intimidating to people who are unfamiliar with financial products?"
Share	"What are your main motivations for going out to vote on election day?"

Table 1: We coded guiding questions into four types based on what the feedback provider is asked to do.

	General Insight	Specific Insight
General Target	"I think the idea is good!"	"Yes but you might want to make it easily skippable."
Specific Target	"I think CTA will really help in your case."	"I think the visual style being modern and playful will make it feel fun."

Table 2: Analyzing Specificity: Comments were coded for whether they addressed a specific target and whether they offered or asked for a specific insight.

Brainstorm questions ask feedback providers to come up with new ideas; critique questions ask feedback providers to evaluate aspects of the game; improve questions ask feedback providers to change existing game elements; share questions ask feedback providers to provide information or opinions for research and user data. We believe these four question categories can be used to code questions beyond the scope of educational game design.

Analyzing Feedback: Specific, Critical, Actionable. A second coding scheme focused on content analysis of the feedback. We developed independent coding schemes to address three key qualities of feedback: specificity (the focus of comments), criticality (sentiment), and actionability (grammatical form).

Specificity. We coded specificity of the target and specificity of the insight. We defined the target as the primary game design element mentioned in the comment. We defined the insight as the interpretation or judgment of the target or other extraneous details (see Table 2).

Two members of the research team reached agreement of 86% for target specificity and 84% for insight specificity when coding comments by practicing coding on an old data set and iterating on a codebook. This codebook defined and classified what components of a comment would be general or specific in terms of its target and insight. In our codebook, we stated that targets such as the overall "game," "music," and "environment" are general, while "objective," "mechanics," and named game elements (e.g. the bird, the tree, the main character) are specific. We also noted that an insight such as "looks good" is general, while "feels natural playing" is specific.

In the first comment in Table 2, for instance, the target is "idea" and the insight is that it "is good." Both the target and the insight are general as they do not delve into the intricacies of any particular game design element, nor do they give any developed interpretation of that element besides stating that it is positive. In the second comment in Table 2, for example,

the target is "CTA" (cognitive task analysis). However, the insight does not elaborate much further on the target, such as explaining how or why the target will "help in your case."

Sentiment. We also coded feedback comments for their sentiment as Praise, Criticism, Neutral, or Both praise and criticism. Two members of the research team were trained using the new coding scheme on an independent training dataset and reached 84% agreement. In our codebook, we defined Praise as feedback that expressed a positive value judgment. Praise could be defined both in terms of the work or the person presenting. However, the feedback generated in this study was overwhelmingly in praise of work rather than people. We defined Criticism as feedback that expressed a negative value judgment. Similar to Praise, Criticism could be defined both in terms of the work or the person presenting, and most of the feedback was critical of work instead of people. Critical comments most often suggested changes to existing game elements. We defined Neutral as feedback that did not express a value judgment or feedback where the value was unclear. Essentially, Neutral feedback was neither praise nor criticism. Neutral comments included suggestions to try something new that did not explicitly critique the current work. Lastly, we defined Both as feedback that contained both positive and negative value judgments. Table 3 gives examples of feedback with different sentiments.

Grammatical Form. Lastly, we coded feedback comments based on their grammatical form as being either Actionable, Descriptive, or a Question. Two members of the research

Grammatical Form and Examples

Actionable	"Have a tutorial level/example round players can go through to see how the mechanics work"
Descriptive	"Immediately this feels it could be great for early elementary kids"
Question	"Are your goals just about cooking pasta or can these translate to cooking other things?"

Table 4: Analyzing Actionability: Comments were coded based on their form: actionable, descriptive, or question.

team were trained using the new coding scheme on an independent training dataset and reached 91% agreement. In our codebook, we defined Actionable as feedback that offered a concrete suggestion for how to fix a problem, called for change, or gave instructions. We defined Descriptive as feedback that either stated an observation or described a problem, feature, or reaction. Lastly, we defined Question as feedback that asked for more clarification. However, rhetorical questions that were also actionable in nature were counted as Actionable instead of as a Question, such as "Perhaps begin with an on boarding process? With a quick tutorial and a practice round?" Table 4 provides examples for each of these three distinct comment forms.

4 RESULTS

What types of guiding questions do students write?

As described in our coding scheme, we identified four types of questions feedback receivers might ask: brainstorm, share, critique, and improve. Feedback receivers wrote Brainstorm and Critique questions more often than Share and Improve questions (see Table 5).

Types of Sentiment and Examples

Praise	"I like the third because it creates the most player dynamic"
Criticism	"The elements of using a telescope could be made more apparent in the design. Right now it seems to be mainly about navigation."
Neutral	"Incorporate worked examples in the game instructions."
Both	"Still seems pretty complex and overwhelming for someone who has no idea about design, but it's definitely better than last time."

Table 3: Analyzing Sentiment: Comments were coded for the sentiment they expressed: praise, criticism, neutral, or both praise and criticism.

Question Type	Frequency
Brainstorm	39%
Critique	38%
Share	13%
Improve	10%

Table 5: Students wrote Brainstorm and Critique questions more often than Share and Improve questions.

We also examined how other factors, such as the timing of question writing, student values, and the question-writing process (individual or collaborative), may have impacted the types of questions feedback receivers wrote. For timing, we looked at how question type varied at different phases of the project. For values, we looked at student ratings of different types of questions or feedback, since what you value can influence what you write. For process, we compared both student values and the types of questions written by teams working individually or collaboratively to write their questions.

The timing of question writing may affect the type of questions teams write. Feedback receivers wrote different types of questions during different phases of the project (see Figure 2). The questions written by each team potentially reflected their particular needs at that point of time in the project. Critique and brainstorm questions were consistently the most common type of question across all three rounds. Most teams asked critique questions during the pitch and interim 2 rounds while most teams asked brainstorm questions during the interim 1 round. This variation across rounds may reflect teams' shift in focus over time.

Students value different types of guiding questions than the instructor. When rating the questions written by their team, feedback receivers rated improve and share questions as more helpful than brainstorm or critique questions, as shown in Figure 3. The instructor believed the value of question types was related to the design cycle. In the debrief interview, the instructor said he believed critique and improve questions would be useful to students throughout their design process, but share and brainstorm questions might only be useful in the beginning of the design process. He felt that once a team had committed to a design, it would be difficult to incorporate feedback about completely new ideas. The instructor did not
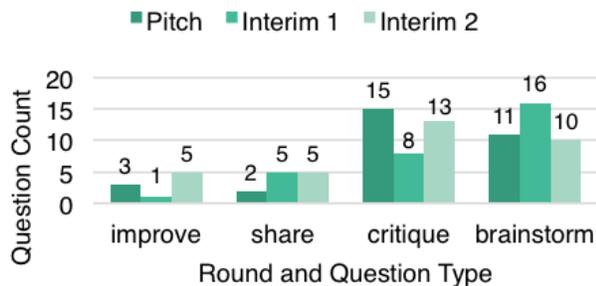
Figure 2: Questions that asked feedback providers to critique were the most common during the pitch and interim 2 rounds. Improve questions were the least common.
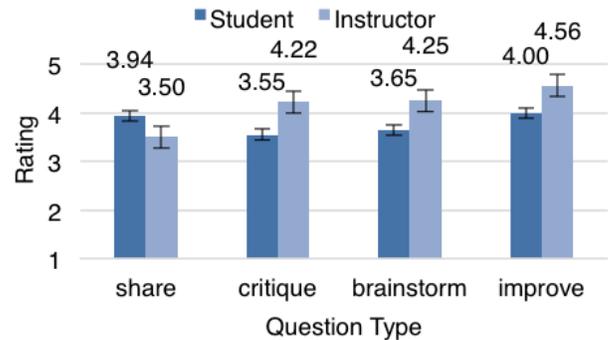
Figure 3: Students rated share and improve questions as more helpful than brainstorm and critique questions. The instructor rated share questions lower than all other question types.

value share questions as much as the other types of questions; he rated improve, brainstorm, and critique questions almost equally, as shown in Figure 3.

Collaborative teams recognized the value of improve questions over share questions. Whether teams wrote questions individually or collaboratively affected how they rated their questions. Collaborative teams rated their improve questions as the most helpful, with a rating of 4.07 ± 0.094. Individual teams rated their share questions as the most helpful, with a rating of 3.94 ± 0.129.

These values may have impacted the kinds of questions teams wrote. All teams wrote more critique and brainstorm questions than improve and share questions, but there were differences between individual and collaborative teams in their use of improve and share questions. Individual teams wrote share questions 21% of the time, roughly three times more often than collaborative teams, who wrote share questions only 7% of the time. Individual teams wrote improve questions only 5% of the time, while collaborative teams wrote improve questions 13% of the time, nearly three times more often.

(How) do students and instructors respond differently?

Instructors gave more specific feedback than students. Instructors and students most often provide feedback with specific target and specific insights (see Figure 4). Differences in feedback specificity emerge when analyzing the second-most common type of feedback. Students' feedback often referred to a specific target (69% of the time), but students were less likely to provide a specific insight about that target. In contrast, instructors almost always provided a specific insight when giving feedback (80% of the time), regardless
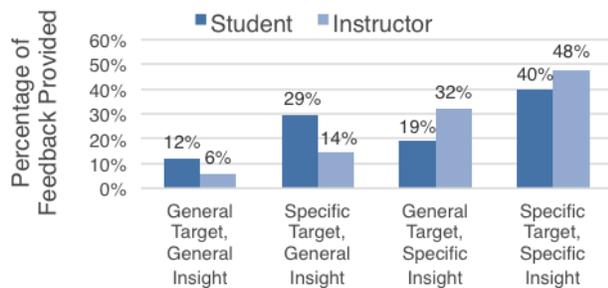
Figure 4: Students gave feedback that focused on specific targets while instructors gave feedback that focused on specific insights.

Feedback Form	Students	Instructor
Actionable	3.65 ± 0.038	3.73 ± 0.107
Descriptive	3.33 ± 0.048	3.19 ± 0.174
Feedback Sentiment		
Criticism	3.64 ± 0.054	3.51 ± 0.164
Neutral	3.47 ± 0.121	3.59 ± 0.120
Praise	3.25 ± 0.121	3.22 ± 0.434

Table 7: When rating feedback, students and instructors both valued actionable and critical feedback.

of whether the target was specific or general. Both for students and the instructor, feedback with a general target and general insight was the least common type of feedback.

Instructors and students were equally critical. Students gave a similar amount of praise and criticism compared to the instructor. We coded feedback sentiment into four categories: praise, criticism, neutral, and both praise and criticism. Across all 29 presentations over three rounds, there were 1,569 comments made by students and 153 comments made by the instructor. Proportionally, students gave neutral and critical feedback nearly as often as the instructor (see Table 6). Students gave more than three times the amount of criticism as praise and more than seven times the amount of neutral feedback as praise.

Instructors gave more actionable feedback than students. We coded feedback into three different types: actionable, descriptive, and question. We found that 50% of student comments were actionable, 44% were descriptive, and 6% were questions. 65% of instructor comments were actionable, 26% were descriptive, and 9% were questions. While students wrote a

similar number of actionable and descriptive comments, instructors emphasized actionable over descriptive comments when giving feedback.
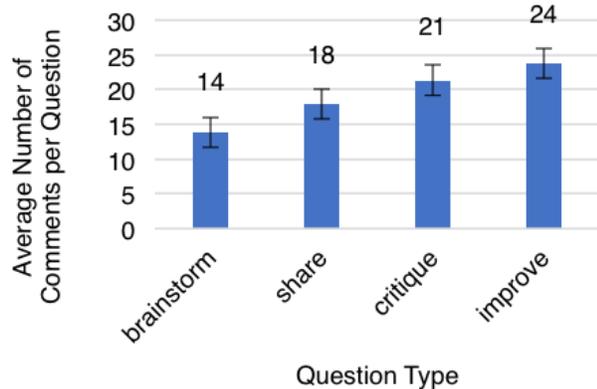
Students and instructors value similar feedback qualities. When asked to rate the feedback, students and instructors rated feedback similarly (see Table 7). Students and instructors both valued actionable feedback more than descriptive feedback. Students and instructors both valued praise less than criticism, although the instructor sample of ratings for praise comments was small. The instructor valued neutral feedback as much as criticism, while students preferred criticism to neutral feedback.

How do the guiding questions affect the feedback?

Improve-type questions yielded more feedback. Questions asking for improvement elicited more feedback than other question types (see Figure 5). Brainstorm questions elicited on average 14 ± 2.163 comments per question while improve questions elicited 24 ± 2.163 comments per question on average. However, each question type is only marginally different than the next, as evidenced by the overlapping confidence intervals in Figure 5.

Improve and Share questions produced more specific feedback than other question types. Improve questions always generated feedback with specific insights. Share and improve questions also had the highest percentage of specific comments (76%) compared to 55% for brainstorm questions and 41% for critique questions. Critique questions elicited the highest percentage of feedback with a general target and general insight (16%) compared to less than 5% for all other question types.

Improve and Critique questions produced more critical feedback than other question types. Improve, share, brainstorm, and critique questions tended to produce either neutral or

Sentiment	Students	Instructor
Neutral	60%	57%
Criticism	27%	29%
Praise	8%	8%
Both	4%	6%

Table 6: Students and instructors were equally critical when providing feedback.

Figure 5: Different types of questions produced different amounts of feedback.

critical feedback. However, the proportion of critical to neutral comments varied greatly. 36% of feedback generated by improve questions was critical and 51% was neutral. Similarly, 31% of feedback generated by critique questions was critical and 48% was neutral. On the other hand, 18% of feedback generated by share questions was critical and 80% was neutral. Also, 22% of feedback generated by brainstorm questions was critical and 73% was neutral. The range between critical and neutral comments for share and brainstorm questions was large relative to the range for improve and critique questions.

Improve and Brainstorm questions produced more actionable feedback than other question types. Improve and brainstorm questions usually produce actionable feedback, share questions usually produce descriptive feedback, and critique questions produce both descriptive and actionable feedback, but mostly descriptive. Brainstorm questions elicited actionable comments 79% of the time and improve questions elicited actionable comments 75% of the time. Share questions elicited descriptive comments 87% of the time. Critique questions elicited descriptive comments 54% of the time and actionable comments 37% of the time.

5 DISCUSSION

Writing student-authored guiding questions presents an opportunity to add constructive and interactive learning activities into the peer feedback process. Feedback receivers wrote four types of guiding questions: brainstorm, critique, share, and improve. Students rated improve and share questions as more helpful than brainstorm and critique questions, even though they wrote more brainstorm and critique questions.

When providing feedback, the instructor was more specific and more actionable than students, but students and the instructor were equally critical. Question type impacted all three factors of feedback quality - specific, critical, and actionable.

Students who collaboratively wrote questions internalized instructor values and standards.

The instructor did not value share questions as much as the other question types, and our analysis showed that share questions did not produce as much actionable or critical feedback. Teams that wrote questions independently valued and wrote improve questions, which produced higher-quality feedback, less than share questions, which produced lower-quality feedback. In contrast, collaborative teams rated their improve questions as most valuable and wrote improve questions more often than share questions.

Collaborative teams more closely aligned with the instructor values of writing questions that elicited critical and actionable feedback. Internalizing instructor values and evaluation criteria is an essential learning aspect of peer feedback [28]. This aligns with ICAP theory, which suggests that the interactive activity that occurred in collaborative teams, but not in individual teams, fostered deeper learning.

Our observations of one collaborative team's question-writing process offer specific examples of how the interaction may have supported feedback receivers in learning and internalizing instructor standards. In the collaborative team question writing session that we observed, students made explicit comments about whether a particular question would be effective, such as "I feel like 'Do you have any ideas...' is a bad question". In this case, the student felt a brainstorm question would not have helped her team, so they discussed what might make a question better. Students in this team discussion also considered whether particular types of feedback would be better obtained by asking their peers in class or running another playtest, which is an expert-level reflection task [11]. Students who wrote questions individually did not have any opportunity for this type of reflection about feedback sources or questions. This finding, supported by ICAP theory, suggests that the interactive learning that occurs in collaborative team discussions about question writing may have positively impacted student's ability to recognize effective questions.

Students and instructors are specific in different ways.

Students focused on a specific target when giving feedback; they would refer to a specific game element but provide only general judgments about that game element. In contrast, instructors focused on specific insights; they would provide specific judgments about either a specific target in the game or the game as a whole. It's easier for students to recognize a specific target (e.g. that character, that music, that object) in

the game and say if it's generally good or generally bad, than it is for them to specifically critique the game as a whole, or specifically critique the target they identified.

When the revised version of Bloom's Taxonomy [1] is applied to these two types of critique, it sheds light on why students and instructors might have commented differently. The students' focus on generally describing a specific target falls under the "understand" category. Students can identify and describe ideas and concepts in the game. On the other hand, the instructor's focus on providing specific insights falls under the "evaluate" category, as they are truly critiquing the game. Evaluation tasks are much more complex than understanding tasks according to the taxonomy, so it makes sense that students struggle to reach that level given their more limited expertise.

Students valued specific feedback while instructors valued actionable feedback.

Students rated share questions as high as improve questions. This may have happened because share questions produced more feedback with both a specific target and specific insight than any other question type. However, the instructor rated share questions as the least valuable, perhaps because share questions produced mostly descriptive feedback. In the debrief interview with the instructor, he said he believed feedback in response to share questions would be "several steps removed from their projects" - in other words, not actionable feedback. When considering the value of questions, students attended to the specificity of feedback while instructors attended to actionable feedback.

Question type influenced the quantity and quality of feedback.

Feedback quantity was influenced by question type. Improve questions had the highest response rate per question, while brainstorm questions had the lowest response rate per question. This makes sense when considering the revised Bloom's Taxonomy, as improve questions most often asked feedback providers to perform Analysis or Evaluation tasks, while brainstorm questions asked for the more complex task of Creation [1].

Interestingly, improve questions were also rated as the most valuable by both students and instructors. This indicates that it may be both easier for feedback providers and more fruitful for feedback receivers to ask questions that require cognitive tasks with mid-level complexity, instead of the high-level creation required for brainstorming.

All three factors of feedback quality - specific, critical, actionable - were influenced by question type. Question type influenced feedback specificity. Share questions produced more feedback with both a specific target and specific insight than any other question type. Improve questions produced

feedback with specific insights exclusively. Question type influenced feedback sentiment. Share and brainstorm questions tended to produce much less critical than neutral feedback when compared to critique or improve questions. Question type also influenced the amount of actionable feedback. Improve questions produced mostly actionable feedback. In contrast, share questions produced mostly descriptive feedback. This indicates that encouraging feedback receivers to write more improve questions and fewer share questions could positively impact the amount of actionable feedback they receive, which might also impact how useful they find the peer feedback exchange.

6 LIMITATIONS

This study was conducted in a specific setting with a specific type of student. The study lasted only 6 weeks of a semester-long course. Students were already accustomed to project-based learning and had already completed design-based assignments prior to our intervention. The course was a game design elective taken by master's students in a competitive educational technology program, so students may have been both more motivated to iterate their work and more capable of improving their work than other students in other settings.

7 IMPLICATIONS AND FUTURE WORK

This study investigates how interactive learning activities before the peer feedback exchange impact peer feedback. But the peer feedback process has several phases: before, during and after feedback exchange. Future studies could seek to understand how interactive learning functions at different phases of the feedback process. For example, what happens when feedback receivers constructively or interactively reflect on the feedback they receive? This deeper understanding of all phases of the peer feedback process can be taken into account when building intellectual models of peer feedback.

A second research implication of this work is that not all guiding questions are equally effective. For example, we found that improve questions elicited better feedback than share questions. Our findings suggest that encouraging students to write more improve questions and fewer share questions would improve the quality of feedback they receive. While this study describes the differences between questions types, it does not identify the underlying cause. Future studies could explore why improve questions produce better feedback than share questions and potentially identify other question types that didn't appear in our sample but would be effective for scaffolding peer feedback exchange.

This work also offers design implications for technology that supports peer feedback. Future systems could not only provide feedback receivers with the opportunity to write

guiding questions, but could also help scaffold the question-writing process to encourage certain question types. For example, improve questions generated better feedback than other question types, but students generated them least frequently. Systems might support feedback receivers in writing more questions of this type.

Future systems could also provide data visualizations to students and instructors to increase their understanding of particular aspects of the peer feedback exchange. The system could provide feedback receivers with data about the success of their questions, such as visualizing data on how many peers responded to each question. Provided with a large enough corpus of student feedback, future systems could also help students understand the quality of feedback elicited by different questions. Digital peer feedback systems could provide the instructor with data about what kinds of questions feedback receivers are writing, which students are struggling to write or answer questions, or even an overview of the entire class's feedback exchange.

Finally, guiding questions could be incorporated into existing or novel systems that support other feedback contexts besides college classrooms, such as peer feedback practices in industry.

8 CONCLUSION

This paper presents the results of a study in a college educational game design course that analyzed how guiding questions affect peer feedback exchange. Applying the lens of ICAP theory, we compared a constructive (individual) question-writing process to an interactive (group) question-writing process. We found that the question writing process influenced the type of questions written and that the questions influenced the quality of feedback received. We also found that students in the interactive condition more closely aligned with the instructors when looking at how they valued different types of questions. The quality of feedback provided in both conditions was higher than that of previous studies that did not use guiding questions. In conclusion, we believe student-authored guiding questions may positively impact the peer feedback process and should be explored further.

REFERENCES

[1] Lorin W Anderson, David R Krathwohl, Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, Paul R Pintrich, James Raths, and Merlin C Wittrock. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition.* Technical Report. https://kkm39ysrp06.storage.googleapis.com/MDgwMTMxOTAzWA==06.pdf

[2] Heidi Andrade and Ying Du. 2005. Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation* (2005). http://www.pareonline.net/pdf/v10n3.pdf

[3] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual Design: Defining Customer-Centered Systems.* Elsevier. 496 pages. https://books.google.com/books?hl=en{&}lr={&}id=JxQaQgOONGIC{&}pgis=1

[4] Judy Bolen. 2009. *An investigation of limited professional development on teacher questioning and learner responses.* Ph.D. Dissertation. Walden University. http://search.proquest.com/openview/df8d581b39771450c81c457e018d47ee/1?pq-origsite=gscholar{&}cbl=18750{&}diss=y

[5] David Boud and Nancy Falchikov. 1989. Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education* 18, 5 (1989), 529–549. https://doi.org/10.1007/BF00138746

[6] Deborah L. Butler and Phillip H. Winne. 1995. Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research* (1995). http://rer.sagepub.com/content/65/3/245.short

[7] MTH Chi and R Wylie. 2014. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist* (2014). http://www.tandfonline.com/doi/abs/10.1080/00461520.2014.965823

[8] Yu-Hui Ching and Yu-Chang Hsu. 2013. Peer feedback to facilitate project-based learning in an online environment. *International Review of Research in Open and Distance Learning* 14, 5 (2013).

[9] Kwangsu Cho and Christian D. Schunn. 2007. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education* 48, 3 (apr 2007), 409–426. https://doi.org/10.1016/j.compedu.2005.02.004

[10] Peggy A. Ertmer, Jennifer C. Richardson, Brian Belland, Denise Camin, Patrick Connolly, Glen Coulthard, Kimfong Lei, and Christopher Mong. 2007. Using Peer Feedback to Enhance the Quality of Student Online Postings: An Exploratory Study. *Journal of Computer-Mediated Communication* 12, 2 (jan 2007), 412–433. https://doi.org/10.1111/j.1083-6101.2007.00331.x

[11] Eureka Foong, Darren Gergle, and Elizabeth M. Gerber. 2017. Novice and Expert Sensemaking of Crowdsourced Design Feedback. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (dec 2017), 1–18. https://doi.org/10.1145/3134680

[12] Charles G. Galloway and Norma I. Mickelson. 1973. Improving Teachers' Questions. *The Elementary School Journal* 74, 3 (dec 1973), 145–148. https://doi.org/10.1086/460813

[13] Graham Gibbs and Claire Simpson. 2004. Conditions under which assessment supports students' learning. *Learning and teaching in higher education* (2004). http://www2.glos.ac.uk/offload/tli/lets/lathe/issue1/issue1.pdf{#}page=5

[14] Sarah Gielen, Elien Peeters, Filip Dochy, Patrick Onghena, and Katrien Struyven. 2010. Improving the effectiveness of peer feedback for learning. *Learning and Instruction* 20, 4 (aug 2010), 304–315. https://doi.org/10.1016/j.learninstruc.2009.08.007

[15] Gunnar Harboe and Elaine M. Huang. 2015. Real-World Affinity Diagramming Practices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15.* ACM Press, New York, New York, USA, 95–104. https://doi.org/10.1145/2702123.2702561

[16] Tasos Hovardas, Olia E. Tsivitanidou, and Zacharias C. Zacharia. 2014. Peer versus expert feedback: An investigation of the quality of peer

feedback among secondary school students. *Computers & Education* 71 (feb 2014), 133–152. https://doi.org/10.1016/J.COMPEDU.2013.09.019

[17] Chinmay Kulkarni, Michael S. Bernstein, and Scott Klemmer. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. *Proceedings from The Second (2015) ACM Conference on Learning @ Scale* (2015), 75–84. https://hci.stanford.edu/publications/2015/PeerStudio/Peerstudio.pdf

[18] Lan Li, Xiongyi Liu, and Allen L. Steckelberg. 2010. Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology* 41, 3 (may 2010), 525–536. https://doi.org/10.1111/j.1467-8535.2009.00968.x

[19] Ngar-Fun Liu and David Carless. 2006. Peer feedback: the learning element of peer assessment. *Teaching in Higher education* (2006). http://www.tandfonline.com/doi/abs/10.1080/13562510600680582

[20] David J. Nicol and Deborah Macfarlane-Dick. 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* (2006). http://ewds.strath.ac.uk/REAP/public/Papers/DN{_}SHE{_}Final.pdf

[21] Alandeom W. Oliveira. 2010. Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching* 47, 4 (apr 2010), 422–453. https://doi.org/10.1002/tea.20345

[22] D. Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18, 2 (jun 1989), 119–144. https://doi.org/10.1007/BF00117714

[23] Amy Shannon, Jessica Hammer, Hassler Thurston, Natalie Diehl, and Steven P. Dow. 2016. PeerPresents: A Web-Based System for In-Class Peer Feedback during Student Presentations. In *Designing Interactive Systems*. 447–458.

[24] Amy Shannon, Alex Sciuto, Danielle Hu, Steven P. Dow, and Jessica Hammer. 2017. Better Organization or a Source of Distraction? Introducing Digital Peer Feedback to a Paper-Based Classroom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*. ACM. https://doi.org/10.1145/3025453.3025564

[25] David Tinapple, Loren Olson, and John Sadauskas. 2013. CritViz: Web-based software supporting peer critique in large creative classrooms. *Bulletin of the IEEE Technical Committee on Learning Technology* (2013). http://www.ieeetclt.org/issues/january2013/Tinapple.pdf

[26] Kieth Topping. 1998. Peer assessment between students in colleges and universities. *Review of educational Research* (1998). http://rer.sagepub.com/content/68/3/249.short

[27] A Yuan, K Luther, and M Krause. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowd-sourced Design Critiques. *(CSCW'16) Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (2016), 1005–1017. https://www.kurtluther.com/pdf/CrowdCrit{_}CSCW{_}2016{_}camera.pdf

[28] Haiyi Zhu, Steven P. Dow, Robert E. Kraut, and Aniket Kittur. 2014. Reviewing versus doing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*. ACM Press, New York, New York, USA, 1445–1455. https://doi.org/10.1145/2531602.2531718