Designing Interactive Scaffolds to Encourage Reflection on Peer Feedback

Amy Cook
University of Memphis
Memphis, USA
ascook@memphis.edu

Steven Dow
Univ. of California at San Diego
San Diego, USA
spdow@ucsd.edu

Jessica Hammer
Carnegie Mellon University
Pittsburgh, USA
hammerj@andrew.cmu.edu

ABSTRACT
Feedback is a key element of project-based learning, but only if students reflect on and learn from the feedback they receive. Students often struggle to deeply engage with feedback, whether due to lack of confidence, time, or skill. This work seeks to identify challenges that make reflecting on feedback difficult for students, and to design possible solutions for supporting reflection. Through observing two university game design courses, our research found that without concrete reflection strategies, students tended to be attracted to feedback that looks useful, but does not necessarily them move forward. When we introduced three different reflection scaffolds to support students, we found that the most effective approach promoted interactive learning by allowing time for self-reflection before team reflection, offering time limits, providing activities for feedback prioritization, helping teams align their goals, and equalizing team member participation. We present design guidelines for future systems to support reflection on feedback.

Author Keywords
Peer feedback; reflection; education technology.

CSS Concepts
• **Applied computing~Collaborative learning**

INTRODUCTION
In project-based learning, feedback can be useful and insightful, but only if students genuinely reflect on the feedback [11]. Strategies and platforms for exchanging peer and crowd feedback have helped instructors ensure that students get a large quantity of feedback, in greater diversity, and in a shorter amount of time (e.g. [15]). However, students often have difficultly reflecting on feedback and sometimes don't reflect at all [15]. Peer feedback varies in quality [19], and typically exceeds the quantity of feedback provided by an expert or instructor. The sheer quantity and diversity of perspectives enabled by new feedback tools has potentially made it more difficult for students to thoughtfully reflect on feedback. While many researchers focus on how to improve feedback quality [4] and digital tools have made feedback easier to gather—whether it's from peers, experts, or crowds (e.g. [15, 28, 30])— students typically are not trained to make sense of feedback that is variable, copious, and diverse.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

DIS '20, July 6–10, 2020, Eindhoven, Netherlands
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-6974-9/20/07…$15.00
https://doi.org/10.1145/3357236.3395480

Students need support, both for organizing feedback provided by peers and for deciding how to move forward. One strategy for supporting students in project-based learning environments is to provide scaffolding to guide students through a problem-solving procedure [22, 29]. In order to design effective procedures and scaffolds for reflecting on peer feedback, we need to identify the challenges students face during reflection and unpack underlying factors contributing to these challenges.

We observed students in two university game design courses engaging with peer feedback during their weekly team meetings. We chose game design because it is representative of project-based teamwork, students have diverse skill sets (artists, programmers, game designers), and the instructors considered peer feedback an essential learning goal for the course. We designed three scaffolds for supporting reflection and analyzed how students responded to those possible solutions. We found that students preferred the protocol that scaffolds self-reflection before team sharing, and that, in general, the protocols influenced what goals students expressed during reflection. For example, one protocol promoted a focus on sensemaking, as students wanted to understand the feedback they received, while another protocol helped students focus on iteration, as students wanted to use the feedback they received to improve their project. We also found that students have difficulty recognizing high quality feedback (specific, critical and actionable [4, 27])and encounter team dynamic challenges during reflection. We conclude with implications for future systems designed to support interactive learning during reflection on feedback.

BACKGROUND
Reflecting on and integrating feedback into future work is an essential skill, not just for the classroom but also for the jobs students will pursue after graduation. Researchers have found that in the game design industry, reflection on feedback is expected, but it is still not supported [26]. Thus, it is even more important for students to learn reflection skills while in the low-stakes classroom environment.

Reflecting on feedback is an opportunity for learning

Reflection encourages students to engage in purposeful thinking, participate in a cycle of inquiry, and form reasoned judgments around a goal [6, 14, 23]. Reflection supports learning in many domains [20], and offers a critical approach to assimilating feedback and improving solutions in project-based STEM learning [25]. In addition, reflection on peer feedback can help students improve their self-regulated learning skills and help them understand the criteria for success in their discipline [1, 3, 21].

Prior work has shown that it is essential for students to reflect in order to integrate feedback into future work [3, 11, 16]. However, students often find reflection challenging, particularly when asked to engage in reflection over time [31]. Three characteristics of peer feedback – that it is copious, diverse, and of variable quality – introduce difficulty in the reflection process. Students untrained in best reflective practice must deal with a large volume of often contradictory feedback without clear indicators of which feedback is "good" or "bad".

Current support for reflection on feedback is lacking.

Peer feedback systems facilitate feedback exchange on student work [15]. Existing peer feedback systems struggle to support students in two ways: 1) explicit support for reflection is minimal, particularly for feedback receivers, and 2) systems are designed for individual feedback receivers, not for teams.

Support for reflection on feedback received is minimal

Some systems help feedback givers reflect on the quality of feedback they **provide** to others with tools like rubrics or back-feedback [3, 28, 32]. But, most existing peer feedback systems provide minimal, if any, support for reflection on the feedback students **receive**. Prior work reveals three commonly used "minimal supports" for reflection on feedback received.

View Only: Many systems allow students to *view* feedback, without any scaffolding for sensemaking or reflection. While some systems allow students to resubmit assignments for additional feedback [3], they do not include features to help students integrate feedback into their revisions. One tool posted student scores on a leaderboard so students could see how their work compared to classmates' [30], but the tool did not help students understand the reasons for their ranking. Simply viewing feedback does not mean students have engaged in reflection [2].

General Prompt: Rarely, systems will explicitly ask students to reflect by providing a general prompt. For example, one tool provided a text box for students to enter reflections on the feedback they received, but researchers found that students rarely used this feature; only 100 out of 3600 students wrote reflections [15] and the quality or impact of those reflections was not discussed.

Ratings: The most common feature to support reflection is to ask students to rate the quality of the feedback they receive [9, 18]. Students might be given a Likert scale for helpfulness, or a 5-star rating system, for example. While rating feedback is certainly a useful activity, there are limitations to its helpfulness. As the amount of feedback increases, rating becomes time consuming, and it can be difficult to decide what to do next on a project when you have dozens of comments rated 3 or 4 out of 5.

In-depth support such as prompts or scripts, prioritization activities like writing next steps, or providing reflection strategies and protocols were absent in prior work.

Most systems support individual reflection, not teams

Another limiting factor in prior work is that most peer feedback systems are designed for an individual to receive feedback, not for a team to receive feedback [3, 7, 9, 18]. Team dynamic issues are common in project-based learning environments, particularly at key points in the design process like when a team is using feedback to decide what to do next, and students struggle to resolve these issues. We found only two systems that explicitly supported team reflection [8, 28] but in both cases the support was minimal (as per above).

Interactive learning can support reflection.

Chi et al [2] categorized learning activities into four types, from most effective to least effective: interactive, constructive, active, and passive. When applied to reflection activities, this categorization offers insights about how to design effective support for reflection.

For example, passive reflection might involve listening to someone give you feedback, with no further action taken. Active reflection could be listening to someone give you feedback and rating the helpfulness of each feedback comment you heard. Reflecting in a constructive way, which involves the learner constructing new information, might ask a learner to use the feedback they received to write next steps for their project. Interactive reflection, which involves learners constructing new information together, might have students on a project team developing their next steps together through discussion. Chi et al theorize that the interactive activity would foster the deepest learning, as a mutual exchange of ideas can result in new ideas that neither individual knew previously nor could generate alone.

Research Questions

This work seeks to validate the challenges students face during reflection on peer feedback, and investigate potential reflection supports for their helpfulness in addressing student challenges. We explore the following two research questions.

1. Identifying Challenges: What challenges do students encounter when reflecting on feedback?

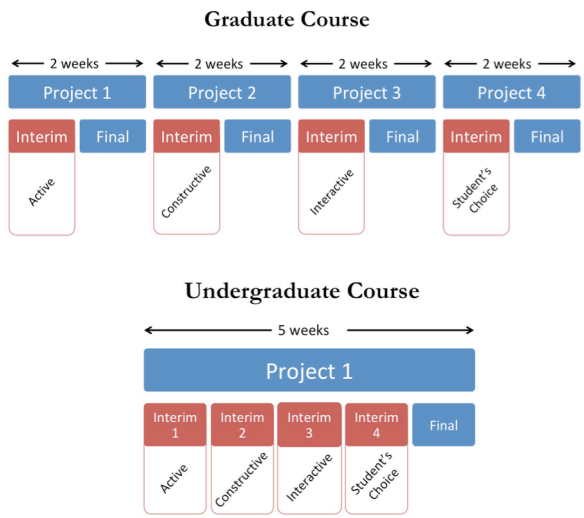

Figure 1: The graduate course had four projects, each lasting two weeks. Students participated in the study every other week, for four total interventions. The undergraduate course had one project that lasted five weeks including the final. Students participated in the study every week, for four total interventions.

2. Evaluating Design of Scaffolds: What kind of scaffolding (active, constructive, or interactive) helps students reflect on feedback?

METHOD

This study seeks to both observe student reflection to identify challenges students face, and design options for reflection support to analyze how students benefit from different types of scaffolding during reflection.

Participants

We conducted this study in two university game design courses. Both courses included students who were divided into teams by the instructor as they completed project-based game design assignments over the course of a semester. Students would present a new version of their game in class every week to receive peer feedback. The larger course included 90 graduate students, divided into 15 teams of 6. In this course, students completed a series of two-week projects and teams were randomized between projects. The smaller course included 15 undergraduate students divided into three teams of 5. In this course, students completed just one project, so teams were consistent throughout the study.

Procedure

Training

Before the study began, a researcher conducted an in-class training session. During this training session, students were asked about their prior experience with peer feedback and what they thought made feedback more or less helpful. Then students were shown research on effective characteristics for feedback: specific, critical, and actionable [21, 24, 32]. Students were shown example feedback comments and given opportunities to revise comments to make them more effective according to the research-backed criteria. Students were then introduced to the study procedures.

Course Context

In both the graduate and undergraduate courses, students participated in the study after each interim presentation of their work in class; because of our focus on formative feedback, we did not conduct the study when students presented a final version of their work. Figure 1 shows how the study fit into the overall courses. In the graduate course, each project lasted two weeks, so students participated in the study every other week for the entire semester. In the undergraduate course, Project 1 lasted 5 weeks so students participated in the study every week for 4 weeks. In both courses, participating in feedback exchange was a requirement for the course. Student teams who chose to participate in the study received their feedback in a reflection meeting with the researcher outside of class, while teams who chose not to participate could pick up their notecards from the researcher at the end of class and meet on their own to reflect.

Weekly Procedure

Students would present their work in class for peer feedback. Before class, teams could sign up for a meeting time. During class, every team presented their work and received feedback. Feedback was written on paper notecards, and feedback providers did not write their name on the notecards. After class, teams would meet with a researcher at their chosen time slot. During this meeting, the researcher would obtain consent, give the team their feedback, guide the team through a reflection protocol, and administer a check-in survey.

Reflection Protocols

We designed three reflection protocols for students to follow based on prior work in iteration-focused reflective feedback practices (see Figure 2). Chi's interactive learning framework served as a brainstorming tool to generate different protocols that were possible to implement in a digital system. We selected one scaffold for each of Chi's three learning styles—active, constructive, or interactive—and wrote a protocol for each type of scaffold. While Chi did inspire other ideas for scaffolds, we chose three for this preliminary investigation. Our protocols built on existing research in effective reflection, such as the use of prompts, scripts, and prioritization activities [3].

At the start of all three conditions, students had 10 minutes to read the feedback their team received. As they read each comment, students rated the helpfulness of the comment on a scale from 1 to 5 (a "minimal" support activity used in existing systems). The activities for the next 15 minutes varied by condition.

SPECIFIC	ACTIONABLE	CRITICAL
Target • Specific • General Insight • Specific • General	• Actionable • Descriptive • Question	• Criticism • Praise • Neutral • Both praise and criticism

Table 1: Researchers coded for three qualities of successful feedback: specific, actionable, and critical.

In the Active condition, after students finished reading and rating, they had 15 minutes to discuss as a team. In the Constructive condition, students had an additional 5 minutes of silence to individually write a list of what they thought were the most important actions to take on this project. After students wrote their list, they had 10 minutes to discuss as a team. In the Interactive condition, students discussed with their team immediately after the reading period. Then they had 5 minutes to collaboratively write a list of next steps that the entire team agreed on. All protocols ended with a 5 minute check-in survey (described below). Students in both courses were able to try all three protocols over the course of their project. In the final reflection meeting, students could choose which protocol they preferred.

The reflection activities were implemented using paper rather than a digital system due to the constraints of the particular courses. However, the findings from this study could be used to inform the future system design of an effective reflection interface.

Voluntary Participation

Teams could decide on each presentation day whether they wanted to participate in the study or not. If at least two team members were able to participate, then the team would sign up for a time to meet with a researcher after their section of class ended. All reflection meetings were conducted on the same day as the in-class feedback exchange. Individuals

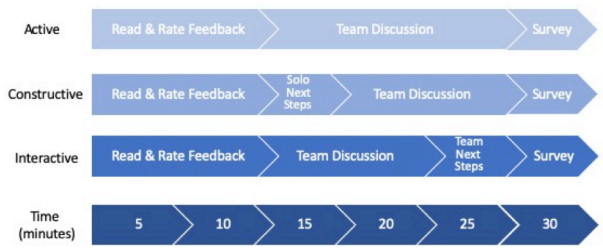

Figure 2: Students were guided through one of three reflection protocols during each reflection meeting. All three protocols took 30 minutes, beginning with a reading period, ending with a 5-minute survey, and including at least 10 minutes for team discussion.

who participated were compensated with a $5 Amazon gift card for each meeting they attended.

Weekly Student Check-In Survey

The check-in survey asked open-ended questions about student likes and dislikes of the reflection protocol they experienced that day, and a series of Likert-scale questions about their team dynamics and their opinions about peer feedback. The survey took less than 5 minutes to complete. The check-in survey was given to all students at the end of every reflection meeting, regardless of which protocol was used. Students answered the survey questions individually on paper before exiting the reflection meeting.

Measures and Analyses

This study seeks to identify challenges students face during reflection and analyze how students respond to different reflection protocols. To achieve these goals, we collected four types of data: quality of peer feedback received, quality of reflection on peer feedback, student preferences among protocols, and team dynamics.

Quality of Peer Feedback

The quality of peer feedback was measured in two ways. First, how often do students provide specific, actionable, or critical feedback? We used a previously developed coding scheme for feedback quality, as published in [4, 27]. Three undergraduate research assistants coded for three qualities of successful peer feedback: specific, actionable, and critical. The research assistants first practiced coding on a training dataset until they reached at least 75% agreement in each coding category, then they coded new data independently. These codes are listed in Table 1, and described in detail in [4].

Second, how do students perceive the peers who provide feedback? We analyzed data from the check-in survey from Likert-scale questions asking students their opinions about the peer feedback process and the peer feedback providers.

Quality of Student Reflection

To measure how well students reflected on the feedback they received, we examined the data through the lens of three questions.

First, how do students rate the feedback they received? Students rated the feedback they received on a scale of 1 to 5. We analyzed average student ratings and whether ratings changed over time.

Second, do student ratings align with expert codes? To answer this question, we compared the researcher codes for the three indicators of feedback quality (see Table 1) to the student ratings of feedback helpfulness on a scale of 1 to 5. Because the researchers coded individual comments, while students rated entire notecards (which might contain 3 comments or more) on a scale of 1 to 5, we calculated a numerical expert rating based on the codes to compare to student ratings. We calculated the numerical rating by first giving each comment a score from 0 to 4, where one point was given for each code (critical, actionable, specific target,

specific insight). We then took the average numerical rating of all comments on the card to create a final score for the entire notecard that could be compared to the student ratings.

Third, do student ratings align with those of other team members? We measured team disagreement by calculating the standard deviation among team member's ratings of individual notecards. We report the average standard deviation across all teams for each interim presentation, and analyze whether ratings change over time.

Student Preferences

We measured student preferences in two ways: students' choices for which protocol to follow and students' expressions of preference. To quantify students' protocol choices, we counted how many teams choose each reflection protocol during the final "choose your own adventure" week. To analyze student preferences, one researcher categorized how students responded to questions about their likes and dislikes for each reflection protocol using a bottom-up coding approach to identify common themes. We did not have the resources to transcribe all our video data of team discussion. We selected quotes from the survey responses that best illustrate each theme.

Team Dynamics

To analyze team dynamics, a team of researchers wrote narrative summaries of the videos of team meetings. We then analyzed the summaries through the lens of one question: What strategies do students use to understand, discuss, and agree about feedback? We also reviewed the open-ended responses to team dynamics questions in the weekly check-in survey.

RESULTS

We collected data from two different game design courses. These courses had a slightly different student population and slightly different project structure. When we analyzed descriptive data from this study, we evaluated whether the two courses were distinctly different.

We compared several measures between the undergrad and graduate course. We found that students in the undergraduate class gave an average of 3.6 feedback comments per presentation (SD = 1.6), and students in the graduate class gave an average of 3.2 feedback comments per presentation (SD = 1.5). Undergraduate feedback was on average 212 characters long (SD = 111), while graduate feedback was on average 181 characters long (SD = 106). Undergraduate comments received an average rating of 2.6 out of 5 from their peers (SD = 1.1), while graduate comments received an average of 2.8 out of 5 from their peers (SD = 1.1). Because we found no differences in these measures, we decided to combine data from both courses for all further analyses presented in this paper.

Identifying Challenges: Why Reflection is Difficult for Students

We identified three major pitfalls students encounter during reflection that contribute to the difficulty of this activity. First, students have trouble recognizing useful feedback. Second, students use reflection strategies that do not help them identify useful feedback. Third, students encounter team dynamics issues while reflecting on feedback.

Students have trouble recognizing useful feedback.

Longer comments get higher ratings.

As shown in Figure 3, comment length was a moderately accurate predictor of student ratings (ratings = 2.03 + .004*numberOfCharacters, R^2 = .299). Longer comments were rated higher by students, even though comment length was not an accurate predictor of expert codes (R^2 = .001). However, there were a few comments that, although short, still received high ratings from students. These comments tended to include a combination of general praise and a concrete suggestion (specific and actionable). For example, "Really cool idea and interaction. Might be better if more character B is inside the world. What if you can sweep the popcorns into the bucket? Audio is great" (P0) and "Very thorough, complete experience. During the early narration, maybe talk a bit about the sun being mean or whatever" (P1) were both highly rated comments with less than 150 characters.

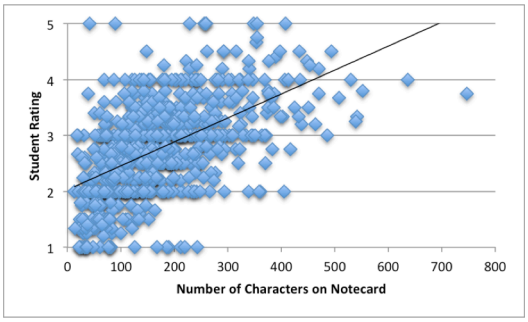

Figure 3: Student ratings aligned with comment length.

Higher quality comments do not get higher ratings.

We used our expert codes to generate a numerical rating for the percentage of specific, critical, or actionable feedback on a notecard, and used linear regression to quantify the strength of the relationship between the variables. Student ratings did not align with the amount of critical feedback (R^2 = .000). Student ratings did not align with the amount of actionable feedback (R^2 = .009). Student ratings also did not align with the amount of specific feedback for target (R^2 = .005) or for insight (R^2 = .017). The lack of relationship between expert ratings and student ratings was true regardless of whether or not we included "fluff" comments (like "Good job!" or "Your music was awesome") that might have a low expert score but could be valued by students as emotional support.

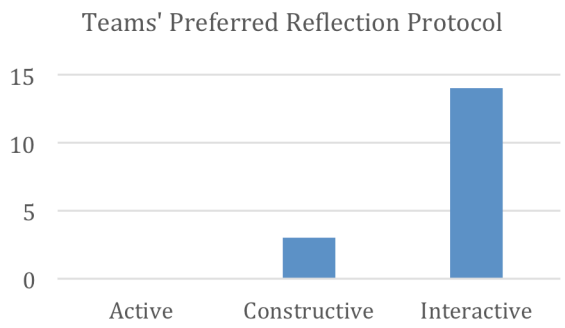

Figure 4: When given a choice, 11 out of 14 teams opted to use the interactive reflection protocol. No teams preferred the Active protocol.

The most frequent peer feedback gets implemented.

Teams often counted how often a particular comment was repeated as a way to decide what to focus on and what to ignore. Team C started their discussion by noting that the most frequent piece of feedback was to shorten the game, and quickly agreed to follow this feedback. Team E also started their discussion with the most frequent feedback. Team D used counting to decide what to ignore – if only one person expressed a minority opinion, they felt comfortable ignoring that disagreement and going with the most popular opinion. While this method could be helpful in identifying the issues the multiple people notice, focusing on only the lowest common denominator feedback might mean that deeper insights get overlooked.

Provider status often trumps critical reflection.

Instructor feedback gets implemented without reflection.

Teams sometimes disagreed about whether to follow peer feedback or instructor feedback. We observed three approaches to resolving the conflict.

In the "worst case" scenario, teams would not acknowledge the contradiction at all. In these team discussions, the instructor feedback was often considered to be more important than peers' opinions, to the point where some team members wanted to ignore peer feedback completely and only consider the instructor comments.

A slightly more reflective approach involved acknowledging the contradiction between peer and instructor feedback, but not resolving the contradiction through reflection. For example, in Team A's meeting, one team member argued for keeping a section of dialogue because the peers really liked it, but another team member argued for removing the dialogue because the instructor didn't like it. The students arguing for the instructor's point of view did not offer any additional reasons for their side of the argument other than because it was the instructor who said so.

In the best-case scenario, a team would be able to reflect on the contradictory feedback and come up with an effective next step for their work. For example, when Team B discussed a transition point in their game, one person noted the negative feedback from peers while another noted the positive feedback from the instructor. In this case, a third team member was able to reflect and suggest that perhaps their transition was conceptually ok, but not visually ok. The team decided to add a tunnel vision effect to this transition point to improve the visual experience.

Teams often implemented leaders' ideas without consensus

Team leadership also affected how students interacted during feedback meetings. One student remarked on the check-in survey "Feedback is all justified. Bad team communication. Dominant producer" (P23). Other students remarked that one team member might take over the reflection meeting rather than soliciting everyone's opinions (for example, "Make the list seemed very one-person led" (P119)).

Students struggle to align with their teams

In the check in surveys, students often mentioned a desire to include the whole team in feedback discussion. However this desire was often expressed after explaining a negative team interaction. Students commented negatively about the team dynamics both when "not all of the team was here" (P118) for the meeting and when individual team members were either "monopolizing" (P122) the discussion or if a "quieter member of the group" was saying nothing at all (P150).

Based on this data, we see that students need help with recognizing useful feedback, critically reflecting on feedback from a high-status source, and promoting positive team dynamics.

Evaluating Design: How Students Respond to Reflection Scaffolds

To understand student reactions to the different reflection protocols, we analyzed student responses to the check-in survey questions and counted how many teams chose each reflection protocol during the "choose your own adventure" week. One difficulty that students face during reflection is using appropriate strategies to critically reflect on feedback. Part of this issue is that students have diverging goals, and therefore disagree on which strategy to use. Therefore when analyzing survey responses, we focused on goals that students identified for their meeting. We also present data on commonly discussed elements of the protocol, such as specific tasks and timing, as that might help future researchers in designing their own protocols.

Students preferred the Interactive protocol

In the "choose your own adventure" week, most teams (11 out of 14) preferred to follow the interactive protocol (see Figure 4). Only three teams chose the constructive protocol, and no teams chose the active protocol, even though the active protocol (which asked students to just read the feedback then talk about it) is what teams told us they would typically do in a reflection meeting without researcher intervention.

Reflection protocol influenced student goals

When commenting on their likes and dislikes in the check-in surveys, students expressed their goals for the reflection process and how the activity did or did not help them achieve these goals. We identified the following five goals using a bottom up coding approach:

- **Sensemaking**. Students wanted to read and understand the feedback they received. Students liked that they "got to all read the feedback" (P2), that reflection "gave us time to process" (P17), and that they "really got to know what's the audience's real feel" (P7).
- **Iteration**. Students wanted to use the feedback they received to iterate their project. They thought of feedback in terms of "different perspectives for what can be done with our project" (P28) and that the reflection "forced me to articulate what it was that I felt the group should move towards goal-wise in the next week" (P61).
- **Prioritization**. Students wanted to know which feedback was the most important, and where to focus next on their project. They used reflection as an opportunity for "reinforcing the highest priorities" (P77), and they appreciated that the reflection method "makes you consciously prioritize which pieces of feedback have the most impact on your situation" (P186).
- **Team building**. Students wanted their team to reach agreement and to avoid negative team dynamics. They said things like, "I like how everyone is on the same page" (P30), and that the reflection meeting "gets everyone in one room" (P47) with time to "engage with other team members" (P51).
- **Efficiency**. Students wanted to feel their time was well spent during meetings and to avoid wasting time. One student said of the reflection meeting "It actually was so much more efficient than just randomly getting together to 'talk'" (P86), or that the thing they liked best about reflection meeting was that it was "very focused, very precise, very efficient" (P147).

The frequency of some of these goals changed depending on which conditions students were in (see Figure 5). As students progressed from active to constructive to interactive, they focused less on just reading and understanding the feedback (sensemaking) and more on using the feedback to improve their work (iteration). Students also revealed a tradeoff – the individual list-making in the constructive activity increased a sense of prioritization but decreased a sense of team building.

Students valued individual reflection time before group discussion

Time to reflect individually before speaking with the group was highly valued and appreciated. Students commented that this time "helped us get our personal thoughts down" (P77), "[gave] individuals more time to think about it before discussing" (P53), and provided time to "come to my own conclusions before sharing" (P94). Students felt this made discussion more efficient because they didn't have to "waste time thinking of what to say next" (P37).

Ratings and next steps helped students prioritize

Students felt the ratings helped them identify what feedback to focus on and prioritize, saying "rating system makes you consciously prioritize which pieces of feedback have the most impact on your situation" (P186). However, they also noted that their own ratings were likely inconsistent over time and would tend to drift towards one end of the spectrum or the other, saying "[I] feel like my ranking of feedback is inconsistent, some days I'm more critical of the feedback author than others" (P157). Students also expressed a feeling of information overload after too many ratings.

When next steps were written individually, students felt the list of next steps helped them prioritize their own thoughts. They used the lists to compare priorities across team members and to identify areas of agreement – "focused next steps allowed us to easily see what we agree on, but perhaps more importantly our priorities" (P38). They felt "compiling a list of things to do made the subsequent discussion more focused/productive" (P54). When written collaboratively, students felt they were able to identify their priorities as a group – "it forced us to prioritize our next steps and turned our goal-oriented items into action-oriented items" (P211). However, one student mentioned the list-making process was "very one-person led" (P119), and another expressed a direct preference for individual list-making (P48).

Students noticed the time constraints

Students had mixed opinions about the time limit researchers placed on the activities. Some students

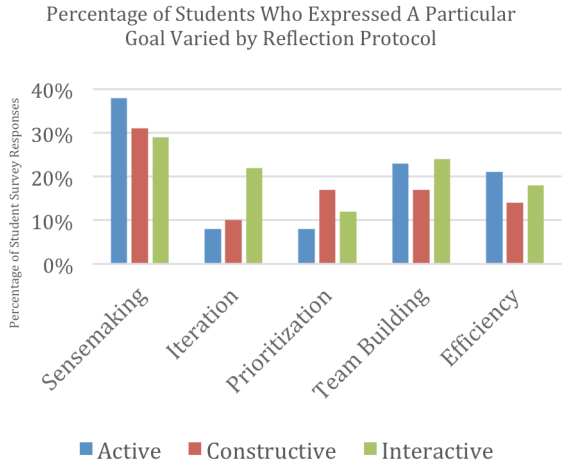

Figure 5: The goals students expressed for the reflection activity changed based on the reflection protocol. Students focused less on sensemaking and more on iteration as the reflection became more interactive. Students also highlighted a tradeoff between prioritization and team building in the constructive condition.

commented that the time limit on group discussion was helpful, saying things like “[I liked t]hat it was timed honestly, it made me prioritize what I'm saying” (P6) and “[the] time limit helps [us] summarize feedback quickly” (P155). Many students commented on the efficiency of the meeting. Others complained that there wasn’t enough time for discussion, saying “the time to discuss is a bit limited” (P184) and “The time limit helps people think and speak their minds quickly, but [I] would've liked it if we weren't stopped when we still had things to say” (P154). A minority of students felt they were wasting time during the reflection meeting, particularly when asked to wait for their teammates to finish reading all the feedback. Some students commented on the overall timing of the reflection meeting happening at a helpful time in the overall design process, saying “I liked that we could meet and discuss right after [class, when] it was fresh in our minds” (P164).

DISCUSSION

Students analyze feedback differently than experts

Peer feedback experts refer to specific, actionable, and critical as markers of high-quality feedback [4, 27]. Students indicated that they did not value criticism or actionable feedback, both by their ratings and by the low amounts of critical and actionable feedback they provided. While students often gave specific feedback, specificity alone was not an accurate predictor of student ratings.

However, students may refer to other markers when analyzing feedback and deciding what feedback to implement. Students rated longer feedback as more valuable than shorter comments. Students prioritized clearly articulated suggestions and instructor feedback during their group discussion, regardless of the quality of those comments. Students indicated that they believed more popular comments were more important. While valuing instructor feedback might be necessary, valuing comments based on length or popularity might not be the most effective reflection strategies.

This data does not explain why students undervalue actionability and criticism. This data might also have missed some additional features that could be influencing student ratings of feedback. Future work could unpack students’ values by going beyond numerical ratings. For example, we could ask students to explain why they rate certain comments higher than others and identify what factors are important to them during feedback exchange. In addition to better understanding student values, this data also suggests a need to help students align with instructor values for high quality feedback. This suggests an avenue for designing future interventions.

Reflection strategies influenced student goals and team dynamics

Students came to the reflection meeting with different values, different goals for reflection, and different strategies for reflecting on feedback. The strategies students used during reflection influenced the goals individual students expressed. For example, when students were in the Active protocol, which focused on reading and discussing, students expressed sensemaking and reflection as their primary goal. When students engaged in the Interactive protocol, which involved making a list of next steps for the project, they valued iteration much higher than in the Active protocol. Students also expressed a tradeoff in goals; the constructive activity resulted in a higher expression of prioritization and a lower expression of teamwork than the other protocols.

This data offers evidence that how students reflect might impact not only what they get out of the reflection activity but also their team dynamics. Because we now know that the strategies students use can affect their perceived value of the reflection activity, it is even more important to investigate what strategies are available to students and which strategies are more or less effective.

Reflection provides an opportunity for team alignment

Typically, reflecting on feedback has been viewed as an activity to promote individual student learning. However, the data in this study suggests that reflection can also be viewed as an opportunity for team building and alignment among team members. Students expressed that they valued group discussion during reflection and that they viewed team building as a goal for the reflection activity. They commented on factors that positively or negatively influenced their team’s ability to align, such as team member attendance and participation. Students also preferred the interactive reflection protocol, which provided the most support for team alignment and had the lowest team disagreement.

This data suggests that reflecting on feedback can be viewed as an opportunity for teams to align, rather than merely a chance to decide what to do next on the project. Future reflection protocols could not only frame reflection as a team activity for students, but also provide strategies for resolving team dynamic issues like disagreement about which next step to take on a project.

Implications for future systems

This study provides suggestions for future technology that might help to better support student reflection on peer feedback.

Promote Interactive Learning

This data suggests that reflection activities that promote interactive learning are effective. Students preferred the interactive learning protocol. Students using the interactive protocol expressed iteration as a goal of reflection more often than students using other protocols. There are certainly other ways to have reflection be an interactive learning activity besides the “write next steps” task used in this study. Future systems could explore other types of interactive reflection activities, including those that involve other roles, such as senior student mentors who provide support and mediation on a backchannel while learners reflect verbally [12].

Reflect Individually before Reflecting as a Team

While team building is an important part of reflection, this data also reveals that students value time to process and reflect on feedback individually before they begin team discussion. Prior work supports that there are benefits to allowing students to disagree before they converge [13]. Future systems can encourage and support time for individual reflection. For example, the PeerPresents system could wait to show team rating visualizations until each team member has rated all the feedback.

Offer Time Limits (maybe)

This study revealed both positive and negative implications of enforcing time limits on discussion. Students liked that the time limits made them prioritize feedback discussion and made their meetings more efficient. However, the strict cut off when the timer went off was sometimes frustrating. Future systems could explore when time limits are most helpful, and when giving teams the option to override a time limit might produce better discussion or more positive experiences during team meetings.

Provide Activities for Prioritization

Students wanted to know where to focus, both in terms of which feedback was most important and in terms of what work to do on their project before the deadline. Students responded positively to both rating feedback and writing next steps for their project. Future systems could explicitly recommend those activities to students during reflection, perhaps building on design recommendations for rating crowd-based feedback such as [10]. This type of activity could help students build their self-regulated learning skills during reflection. Future systems could also explore other ways to support prioritization. For example, the instructor could publicly highlight some of the peer feedback comments that they think are important for a particular team. This type of activity could help students learn to interpret the standards of their discipline. This could also help the instructor have a clear idea of the quality of peer feedback exchange in their course, as well as incentivize students to provide more thoughtful comments.

Help Teams Align Goals

In support of team alignment during reflection, future systems could include more explicit opportunities for team members to identify their agreement or disagreement with the team, perhaps through disagreement highlighting [17]. For example, in the PeerPresents system, a feature could be added to visualize each team member's rating of individual comments compared to the team and feedback could be sorted based on agreement. Comments with high variance, indicating team disagreement, could be highlighted.

Equalize Team Member Participation

Students in this study expressed a desire for reflection to be a team building activity. Students notice when participation isn't equal, and those participation inequalities can cause further challenges in project-based courses. In addition, since the critique process is an essential element in design education [5], students who do not participate are missing a learning opportunity. For example, when "dominant" leadership hijacks design direction, this can discourage participation from other students. Or, when team members become silent or absent, they may lose feelings of ownership in the project, lose learning opportunities, and hurt team dynamics. Future systems could explore potential solutions like highlighting individual voices at certain stages of the reflection process. For example, a system could display each team member's ratings of a comment, similar to how Google Forms displays survey data.

Limitations

One major limitation of the study design is that we used data from two different courses. While we examined multiple factors before deciding to include both datasets, there could have been an undetected distinguishing factor indicating the data should have been considered separately.

A second limitation is that the condition order was not varied. Thus we cannot know if the results were based on a change of condition or a passage of time. Perhaps students simply got better at feedback reflection, and that caused their change in perspective, rather than the condition.

Another limitation related to condition order: it could be that students were more likely to focus on iteration at a later stage of their project, rather than because of the Interactive protocol. Note that while this could be true for the undergraduate students who worked on 1 project for 5 weeks, it is likely not true for the graduate students, as they were in the first week of a project for every intervention (see the course timelines in Figure 1).

Finally, we did not measure or evaluate students' final projects, and we did not analyze if or how much feedback was actually implemented into their final projects. Future work could learn more about best feedback practices through this type of evaluation.

CONCLUSION

We conducted a study with two university game design courses. Through our investigation, we identified three challenges students face when reflecting on peer feedback: students struggle to recognize high quality feedback, students let the feedback provider's status trump critical reflection on feedback, and students encounter team dynamic issues. When we presented students with three reflection protocols to help address these challenges, we found that students preferred the interactive protocol, and that the interactive protocol supported students in focusing on iteration as a goal of the reflection meeting. We also identify design guidelines for future systems to support reflection on peer feedback.

ACKNOWLEDGMENTS

We would like to thank the NSF for funding this work under NSF Grant #1821590, #1122320 and #1821618.

REFERENCES

[1] Butler, D.L. and Winne, P.H. 1995. Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*. (1995).

[2] Chi, M. and Wylie, R. 2014. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*. (2014).

[3] Cho, K. and Schunn, C.D. 2007. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*. 48, 3 (Apr. 2007), 409–426.

[4] Cook, A. et al. 2019. How Guiding Questions Facilitate Feedback Exchange in Project-Based Learning. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* (New York, New York, USA, 2019), 1–12.

[5] Costantino, T. 2015. Lessons from art and design education: The role of in-process critique in the creative inquiry process. *Psychology of Aesthetics, Creativity, and the Arts*. 9, 2 (May 2015), 118–121.

[6] Dewey, J. 1933. *How We Think: A Restatement of the Relations of Reflective Thinking to the Educative Process*. D.C. Heath.

[7] Easterday, M.W. et al. 2014. Computer supported novice group critique. *Proceedings of the 2014 conference on Designing interactive systems - DIS '14* (New York, New York, USA, Jun. 2014), 405–414.

[8] Easterday, M.W. et al. 2017. Designing Crowdcritique Systems for Formative Feedback. *International Journal of Artificial Intelligence in Education*. 27, 3 (Sep. 2017), 623–663.

[9] Foong, E. et al. 2017. Novice and Expert Sensemaking of Crowdsourced Design Feedback. *Proceedings of the ACM on Human-Computer Interaction*. 1, CSCW (Dec. 2017), 1–18.

[10] Frens, J. et al. 2018. Supporting answerers with feedback in social QandA. *Proceedings of the 5th Annual ACM Conference on Learning at Scale, L at S 2018* (Jun. 2018).

[11] Gibbs, G. and Simpson, C. 2004. Conditions under which assessment supports students' learning. *Learning and teaching in higher education*. (2004).

[12] Gray, C.M. and Howard, C. 2015. "Why are they not responding to critique?": A student--centered construction of the crit. *LearnxDesign: The 3rd International Conference for Design Education Researchers and PreK-16 Design Educators* (Chicago, IL, 2015), 1680–1700.

[13] Jehn, K. et al. 1997. To agree or not to agree: The effects of value congruence, individual demographic dissimilarity, and conflict on workgroup outcomes. *International Journal of Conflict Management*. 8, 4 (1997), 287–305.

[14] Kolb, D.A. 1984. *Experiential Learning*. Prentice-Hall.

[15] Kulkarni, C. et al. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. *Proceedings from The Second (2015) ACM Conference on Learning @ Scale*. (2015), 75–84.

[16] Liu, N.-F. and Carless, D. 2006. Peer feedback: the learning element of peer assessment. *Teaching in Higher education*. (2006).

[17] Liu, W. et al. 2018. ConsensUs. *ACM Transactions on Social Computing*. 1, 1 (Jan. 2018), 1–26.

[18] Luther, K. et al. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15* (New York, New York, USA, 2015), 473–485.

[19] McMahon, T. 2010. Peer feedback in an undergraduate programme: using action research to overcome students' reluctance to criticise. *Educational Action Research*. 18, 2 (Jun. 2010), 273–287.

[20] Moreno, R. and Mayer, R.E. 2005. Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *Journal of Educational Psychology*. 97, 1 (2005), 117–128.

[21] Nicol, D.J. and Macfarlane-Dick, D. 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*. (2006).

[22] O'Rourke, E. et al. 2015. A framework for automatically generating interactive instructional scaffolding. *Conference on Human Factors in Computing Systems - Proceedings* (Apr. 2015), 1545–1554.

[23] Quinton, S. and Smallbone, T. 2010. Feeding forward: using feedback to promote student reflection and learning – a teaching model. *Innovations in Education and Teaching International*. 47, 1 (Feb. 2010), 125–135.

[24] Sadler, D.R. 1989. Formative assessment and the design of instructional systems. *Instructional Science*. 18, 2 (Jun. 1989), 119–144.

[25] Sargeant, J.M. et al. 2009. Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education*. 14, 3 (Aug. 2009), 399–410.

[26] Seering, J. et al. 2019. Peer Feedback Processes in the Game Industry. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (New York, NY, USA, Oct. 2019), 427–438.

[27] Shannon, A. et al. 2017. Better Organization or a Source of Distraction? *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17* (New York, New York, USA, 2017), 5545–5555.

[28] Shannon, A. et al. 2016. PeerPresents: A Web-Based System for In-Class Peer Feedback during Student Presentations. *Designing Interactive Systems* (2016), 447–458.

[29] Simons, K.D. and Ertmer, P.A. 2005. Scaffolding Disciplined Inquiry in Problem-Based Environments. *International Journal of Learning*. 12, 6 (2005), 297–305.

[30] Tinapple, D. et al. 2013. CritViz: Web-based software supporting peer critique in large creative classrooms. *Bulletin of the IEEE Technical Committee on Learning Technology*. (2013).

[31] Xie, Y. et al. 2008. The effect of peer feedback for blogging on college students’ reflective learning processes. *The Internet and Higher Education*. 11, 1 (2008), 18–25.

[32] Yuan, A. et al. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. *(CSCW’16) Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. (2016), 1005–1017.