Using Interactive Learning Activities to Address Challenges of Peer Feedback Systems

CMU-HCII-19-101
May 2019

Amy Cook
Human-Computer Interaction Institute
Carnegie Mellon University

Thesis Committee
Jessica Hammer (chair) [HCII & ETC, CMU]
Steven Dow (co-chair) [University of California, San Diego]
Ken Koedinger [HCII, CMU]
Marsha Lovett [Psychology, Eberly Center, CMU]

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract
Effective feedback is a central tenet for project-based learning, but the limits of feedback resources become increasingly evident as class size increases. For example, time demands preclude instructors from providing frequent, detailed feedback for every student in large classes. Instructors often turn to peer feedback systems to provide feedback at scale. However, existing systems struggle to engage students in the peer feedback process, improve feedback quality over time, and support reflection on peer feedback.

In this thesis, I describe my work to use interactive learning techniques to address these challenges of peer feedback systems. Interactive learning techniques ask students to generate novel learning-related materials and to collaboratively engage with them, which optimizes learning. My work identifies opportunities to use interactive learning techniques to improve the peer feedback process. I articulated a theoretical framework for in-class peer feedback activities, developed a novel interaction system for in-class peer feedback, and analyzed its impact on the feedback provided. I examine how interactive learning activities introduced both before and after peer feedback exchange impact peer feedback quality, perceived value of peer feedback, and both student and instructor attitudes towards peer feedback.

Acknowledgments

Thank you all! I have had so much love and support from so many mentors, summer interns, staff, friends, and family. I would not have made it without you! To name just a few from my CMU community…

To my advisors: Jess and Steven. You've both been there through it all, from failed experiments and persistently unpredictable system crashes to paper acceptances and grant awards. Our collaboration has evolved into the dream team of research alliances! I am so grateful for your mentorship, support, and guidance. My rollercoaster journey would have been much less enjoyable without you.

To my committee: Ken and Marsha. This work has benefitted from your mentorship and been bolstered by your excitement. Thank you for your thoughtful suggestions and creative ideas.

To all the OH Lab members, past and present. I have been so lucky to have your support, whether expressed as comments on yet another practice talk, bonding time with board games in the lab, or lunch at "chicken on a stick".

To the SCS Grad Student Musical community. I will never forget the memories made hunting for costumes, engineering creative choreography, or during seemingly endless tech week rehearsals.

Most of all, I would like to thank my family, Mom, Dad, and Matthew, to whom this dissertation is dedicated.

Table of Contents

1 INTRODUCTION

Effective peer feedback is an essential aspect of project-based learning [51]. Peer feedback plays a pivotal role in project-based learning and design education, enabling students to learn from one another and to iterate on and improve their projects [39, 46]. Many researchers have demonstrated the value of peer feedback as a method to provide feedback at scale [13, 35, 60]. For mid-size to large classes, peer feedback can potentially increase the amount of feedback and shorten the amount of time to receive feedback, compared to receiving feedback from the instructor alone (e.g. [36]).

Benefits & Challenges of Existing Peer Feedback Methods

Both giving and receiving peer feedback offer a unique learning opportunity for students to become better acquainted with self-assessment, communication, and project improvement [8, 61]. Receiving peer feedback is a critical component of self-regulated learning [8], and giving peer feedback to others can facilitate learning that transfers to a student's own project [38]. Learning to give and receive feedback helps students build a core skillset for project-based education by helping students develop both feedback skills and domain-specific skills [4, 13].

Peer feedback has been shown to improve domain-specific learning. Peer feedback exchange helps feedback receivers make connections between evaluation rubrics and their own work [66]. The activity also benefits the feedback provider by helping them think critically about the evaluation criteria [17]. Both understanding evaluation criteria and applying those evaluations to projects strengthen students' metacognitive skills in their domain.

Peer feedback also helps students develop feedback skills. Feedback providers can learn what qualities make feedback more or less helpful to the person receiving it [14]. Developing feedback skills such as receptiveness to criticism, resolving contradictory feedback, and using feedback to articulate next steps for a project is essential for students' learning[27, 64].

Instructors increasingly turn to digital peer feedback systems to facilitate feedback exchange for their students [13, 36, 60]. Peer feedback systems allow the number of feedback providers to scale proportionally with the size of the class. However, research also reveals a number of challenges with peer feedback systems.

Challenge 1: Engagement.

Students often struggle to **engage** in the peer feedback process. Peer feedback systems are typically time consuming for students outside of class [17], which contributes to students often begrudging the peer feedback process [36]. Students complain about the time burden of preparing their own project, evaluating and commenting on several other students' work, digesting the feedback they receive on their own project, and then revising their work in time for the next deadline. Also, peers typically provide feedback independently, which misses an opportunity for interaction with other students [12]. If students do not fully engage as they give or receive feedback, they squander their chance to learn from feedback exchange.

Challenge 2: Improvement.

Students don't always learn how to **improve** the quality of the feedback they give. Prior work identifies high quality feedback as specific, critical, and actionable [46, 55, 65]. Peer feedback varies in quality [43], but simply giving feedback over and over again is not enough for feedback quality to improve over time. Peer feedback systems typically do not provide training or support to help students improve the quality of feedback they give. Additionally, peer feedback systems usually give students the perspective of only a few peers [4, 43], which limits the number of diverse perspectives a student is exposed to and limits their opportunity to see examples of what good feedback might look like.

Challenge 3: Reflection.

Students may not know how to effectively **reflect** on the feedback they receive. Especially in larger classes when students receive larger quantities of feedback, reflecting in such a way that the feedback is understood and turned into action items for the project can be so difficult that students often give up on the process. Peer feedback is only useful if it's reflected on, and even existing systems that offer reflection support have trouble helping students integrate feedback into their work [36].

While peer feedback systems can be an effective approach, these challenges highlight issues with **engaging** in the peer feedback process, with **improving** feedback quality, and with **reflecting** on feedback.

Addressing These Challenges

My research explores these three key open challenges by designing learning activities to support in-class peer feedback systems in university game-design courses. My work explores designs for making real-time peer feedback more interactive and evaluates the effects of these designs. The end goal of this work is to develop innovative interaction systems that can improve in-class peer feedback.

My research designs learning activities to support in-class peer feedback systems. The ICAP framework introduced by Chi et al [12] defines four types of learning activities: interactive, constructive, active, and passive. *Interactive* learning techniques ask students to generate novel learning-related materials and to collaborate in engaging with those ideas. Interactive learning techniques optimize student learning because they allow students to engage deeply in activities such as co-creation. Interactive learning was shown to be more effective than constructive, active, or passive learning techniques in the context studied [12]. My research investigates how active, constructive, and interactive learning activities align with peer feedback systems.

I identify three distinct phases of peer feedback exchange. Each phase coincides with specific actions for the students. First, students must prepare before the feedback exchange occurs. Preparation can involve honing both their domain-specific skills as they work on their project, and their feedback-specific skills as they prepare to give feedback to others. Second, students give and receive feedback during the feedback exchange. Third, after the exchange has ended, students reflect on the feedback they have received. In typical peer feedback systems, all three phases occur outside of class. In my research, the feedback exchange (phase 2) happens during class. This change reduces the burden on students and creates an opportunity for real-time student interaction. This shift also highlights that feedback exchange is important enough to devote class time to the activity.

These three phases align with the three challenges identified in prior work. When students are preparing for feedback exchange, they need to improve their feedback skills. During feedback exchange, students need to engage with the feedback. After feedback exchange, students need to reflect. See Table 1. The phases serve as a theoretical framework for identifying opportunities to enact learning interventions that will address each challenge. This dissertation is composed of three studies as outlined below.

When?	Preparing for feedback exchange	During feedback exchange	After feedback exchange
What?	IMPROVE	ENGAGE	REFLECT

Table 1: During feedback exchange, interactive learning activities focus on engaging students in the peer feedback process. When preparing for feedback exchange, activities focus on improving peer feedback quality. After feedback exchange, activities focus on supporting reflection.

RQ1: How can we help students Engage in the peer feedback process? Students who don't see the value in peer feedback exchange are unlikely to benefit from it. Typical peer feedback systems are time consuming for students, requiring time outside of class, and students give peer feedback in isolation, missing an opportunity for interactive learning. Bringing digital peer feedback into the classroom might help students value the peer feedback process.

> **Study 1: In-Class Peer Feedback Systems**. To address these challenges, I oversaw the development of and conducted preliminary evaluations on an in-class peer feedback system called PeerPresents [58, 59]. Study 1 demonstrated that the baseline system performs similarly to paper-based feedback with regard to feedback quantity and quality, while keeping students engaged in the feedback process for longer and producing an activity log that can inform instructors. However, I found that students needed additional guidance for improving feedback quality and deciding how to reflect and act on feedback when iterating their projects.

RQ2: How can we help students Improve the quality of feedback they give? Students often do not give high quality peer feedback, and existing peer feedback systems do not help students improve the quality of feedback over time. Providing students with guidelines for what areas of a project to focus on when providing feedback is helpful for feedback providers [54]. Having feedback receivers author the guiding questions could not only help support feedback providers, but also help feedback receivers thoughtfully consider their feedback.

> **Study 2: Student-Authored Guiding Questions**. To address this challenge, I designed and evaluated a learning activity called guiding questions. As an interactive activity, feedback receivers collaboratively discern what feedback would be most helpful to their project; then they write guiding questions to elicit that feedback from their peers. As a constructive activity, feedback receivers individually consider what feedback they would like, and

individually write questions to elicit that feedback. I evaluated the impact of both individual and collaborative student-authored guiding questions compared to instructor-authored questions on perceived quality of peer feedback exchange from the perspective of feedback receivers, feedback providers, and instructors. I found that the type of guiding question impacted the quality of feedback, and that students in the collaborative condition were able to recognize effective questions in a way that students in the individual condition were not.

RQ3: How can we help students Reflect on peer feedback they receive? Peer feedback is only valuable if students reflect on the feedback they receive. Yet existing peer feedback systems have trouble getting students to reflect. Prior work has shown that action-oriented reflection on feedback, rather than unstructured reflection, can be helpful and motivating for students [64].

> **Study 3: Reflection Protocol**. To address this challenge, I developed three versions of a protocol to guide students in incorporating the feedback they receive. Each version is designed around a different type of learning activity: active, constructive, or interactive. I empirically evaluated these three protocols by analyzing the quality of student reflection, student preferences, and team dynamics.

Table 2 shows how this body of research uses interactive learning activities to address known problems of peer feedback systems for each stakeholder. This table also identifies ideas for additional interactive learning activities that could be studied in future work, but are beyond the scope of this dissertation.

For example, in Study 1, all activities happened during class. Students taking the role of feedback receivers presented their work during class. Then students taking the role of feedback providers wrote peer feedback comments and up-voted comments written by other peers. While students were writing, the instructors modeled proper feedback behavior by providing verbal feedback to the presenting students.

Study 2 followed the same in-class peer feedback process as Study 1, but the presenting team prepared for their presentation differently. Feedback receivers discussed with their team about the state of their project and what feedback would be most helpful to them, then collaboratively wrote guiding questions to help guide the feedback providers. Then during class, students providing feedback answered the questions the presenting team asked, instead of providing unscaffolded peer feedback.

Study 3 again followed the same in-class peer feedback process as Study 1, but the presenting team reflected on the feedback they received after the presentation ended. The feedback receivers followed one of our action-oriented reflection protocols to collaboratively decide how to incorporate feedback into the next iteration of their project.

In addition to the three studies above, I also iteratively developed and piloted a peer feedback skills inventory. This online survey seeks to identify if and how students are using effective feedback skills and strategies. The survey asks students about their opinions of giving and receiving feedback, then presents a scenario where students identify strategies for summarizing feedback, resolving conflicting feedback, and using feedback to identify next steps.

Thesis Statement & Overview

My research demonstrates that understanding peer feedback exchange as an end-to-end process, from preparation through reflection, and incorporating interactive learning activities to support every stage in that process improves our ability to use digital systems to address challenges with engagement, improvement, and reflection. My in-class system directly addresses these challenges in project-based game design classes by moving feedback exchange into the classroom, supporting student-authored guiding questions, and scaffolding reflection.

When?			Preparing for feedback exchange	During feedback exchange	After feedback exchange
What?			IMPROVE	ENGAGE	REFLECT
			How?		
Who?	Students	Feedback Receivers	[Study 2] Write Guiding Questions	[Study 1] Present work	[Study 3] Reflect on Feedback
		Feedback Providers	Discuss highlighted examples	Vote on peer feedback	Reflect on Back-feedback
	Instructors		Highlight feedback examples	Model good feedback behavior	Analyze dashboard data

Table 2: Study 1 addresses helping all three stakeholders engage during feedback exchange. Study 2 addresses helping students improve feedback quality before feedback exchange occurs. Study 3 addresses helping feedback receivers reflect on their feedback after feedback exchange.

During the course of this research, I articulated a theoretical framework for in-class peer feedback activities, developed a novel system for in-class peer feedback, and analyzed its impact on the feedback provided. I examine how interactive learning activities before and after in-class peer feedback exchange impact peer feedback quality, the perceived value of peer feedback, and both student and instructor attitudes towards the peer feedback process.

My work is organized around the following contributions:

(1) A survey of the existing literature on peer feedback systems. In Chapter 2, I begin with a comprehensive review of the existing literature on peer feedback systems, in-class technology, and learning activities to support feedback exchange.

(2) A novel in-class peer feedback system. In Chapter 3, I present the design and initial pilot study of PeerPresents. This system is designed to support open-ended in-class peer feedback on student projects.

(3) A framework for implementing learning activities with an in-class peer feedback system. Also in Chapter 3, the framework of considering each phase (before, during, and after feedback exchange) when implementing an in-class system is explained.

(4) Activities for engaging all stakeholders. In Chapter 4, I describe the validation of the PeerPresents digital system in a study that evaluated engagement of the feedback providers, feedback receivers, and instructors when compared to using paper for feedback exchange.

(5) A learning activity for improvement: student-authored guiding questions. In Chapter 5, I present the design of student-authored guiding questions as an interactive learning activity and provide data from its implementation in a university game design course.

(6) A learning activity for reflection: action-oriented group reflection. In Chapter 6, I present the design of three types of reflection activities based on the ICAP framework (active, constructive, and interactive) and explore data from their implementation in two university game design courses. In Chapter 7, I present a peer feedback skills inventory that identifies and describes reflection strategies students use.

(7) A discussion of the implications of this work for future systems. Finally, in Chapter 8, I conclude with a discussion of key takeaways and potential future work.

This work contributes to learning theory by designing new interactive learning activities and by examining the impact of interactive learning activities in a context that was not studied in the original presentation of the ICAP framework. This work contributes to computer science education by providing a theoretical framework and a digital system that could be used in other project-based courses to improve peer feedback exchange.

2 LITERATURE REVIEW

In this chapter, I describe related work concerning peer feedback, digital peer feedback systems, the remaining challenges with existing systems, and interactive learning activities.

PEER FEEDBACK

Benefits of Peer Feedback

Numerous researchers have studied the benefits of peer feedback for all three parties involved: instructors, feedback receivers, and feedback providers. Peer feedback lowers the burden for instructors to generate feedback for the whole class, and allows students to get a larger quantity of feedback than if the instructor were the only feedback provider [61]. It also allows students to receive feedback faster than if every student was waiting on the instructor. This is particularly important given that a recent study found peer feedback helped students improve their grades, but only if delivered within a timely manner [36].

Students who receive feedback improve their self-regulated learning skills [8] and self-assessment abilities [39] through the process of reflecting on feedback and revising their work. Students who provide feedback learn to recognize what "good" work looks like and to correctly interpret standards or criteria [46]. Peer feedback also helps students improve their self-assessment abilities [39]. Students who provide feedback can learn to focus on students' work and performance, rather than on the students themselves or their personal characteristics [24]. This process also helps students understand the desired criteria (relevant), compare their actual performance with these criteria (critical), and engage in action that closes this gap (actionable) [46, 55].

Student vs. Instructor Feedback

Multiple researchers have shown that novice peer feedback can be as effective as receiving feedback from an expert, such as the instructor. Topping [61] found that peer assessment, where students provide a grade in addition to feedback, can improve student performance at least as much as instructor feedback. Cho and Schunn [13] used their SWoRD system for providing feedback on writing assignments to show that students receiving feedback from a group of novices had greater improvement on their next draft than students receiving feedback from a single expert, perhaps because students consider peer feedback carefully rather than blindly adhering to expert suggestions. Yuan et al [65] found that when students used rubrics to give critiques, their feedback was perceived to be as useful as expert feedback.

While experts have deep domain knowledge, peers have the advantage of sharing the experiences and challenges of the student. Experts often poorly estimate what feedback is most helpful to novices [29], in part because experts tend to reference concepts and information that novices do not yet possess [9]. Peers, on the other hand, share concepts and skills with the student receiving the feedback, as they are at a similar level of expertise [13]. Even if the content of the feedback is less expert, it may be easier for students to process and understand. Peer feedback can therefore supplement expert feedback, as each has different strengths and weaknesses.

Challenges of Current Peer Feedback Methods

Classrooms currently implement peer feedback in a variety of ways, but each of these methods have challenges. Verbal feedback techniques, such as question & answer periods or open design critiques, allow the entire group to hear all the feedback given. However, in many classes, whether due to time constraints or social anxiety, not everyone gets to speak. Verbal feedback can also be easy to forget unless written down or recorded.

Written feedback, such as typing comments into an email or filling out paper evaluation forms, allows every peer to give feedback privately and in parallel, so the time required to give feedback does not need to scale to accommodate larger classes. However, these strategies do not give peers the ability to see feedback already given by other peers. Further, handwritten feedback can sometimes be illegible and difficult to organize.

These challenges, particularly the challenge of scaling feedback in large classes, have led to a rise in digital peer feedback systems.

PEER FEEDBACK SYSTEMS

A number of digital peer feedback systems have emerged to help collect and organize written feedback [13, 36, 60]. The increasing demand for feedback in large classes has encouraged instructors to turn to peer feedback systems to ensure that students get personalized input on their work in a reasonable amount of time [13, 36, 42, 60].

Taxonomy of Existing Peer Feedback Systems

To understand the current state of peer feedback technology, I first investigated the successes and pitfalls of existing tools. I reviewed 32 different tools designed to capture qualitative feedback on student assignments, as well as tools designed to elicit responses from students during class or for general audience interactions (see Table 3).

Peer Feedback Tools	In Class	Rubrics	Anonymous	Voluntary	Voting	Reflection
Audience Response Systems	X			X	X	
CritViz		X	X			
SWoRD, TurnItIn		X	X			
PeerStudio		X	X			X
CSGC	X	X			X	
PeerPresents	X	X	X	X	X	X

Table 3: Features of existing peer feedback tools compared to PeerPresents.

I discovered systems to review by following citation trails, searching academic and non-academic databases, and getting suggestions from experts. These systems supported activities such as peer feedback, student or instructor critique, facilitating discussion, answering questions during lecture, asking questions during lecture, and polling large groups. I compared the systems using a competitive analysis. Rather than presenting all 31 systems, this section summarizes the systems most influential to the design of a system for in-class peer feedback.

Feedback Systems

Most peer feedback systems are specifically designed for students to provide feedback outside of class [13, 32, 36, 60, 67]. In some cases, like TurnItIn [67], the peer feedback system would connect to a school's online learning management system. I found one system designed for use during class [16], but this system—used to encourage novice group critique in a design course—also asked students to provide written feedback before the in-class critique session.

The peer review process is commonly characterized by a required number of reviews for each student [13, 36, 67], structured rubrics authored by the instructor to guide feedback [60], and a revision period before the final deadline [7]. In the case of PeerStudio, which facilitates rapid peer feedback to students in MOOCs, survey responses indicated that many students felt their schedule was too busy to revise. One student complained about the workload, saying that "(the instructors) expect us to read some forty page essays, then write the critiques and then review two other people, and then make changes on our work... twice a week" [36].

While some systems permit resubmission of assignments [13], reflecting on feedback was a task often ignored by peer feedback systems. Some provided only a way for students to view the feedback they received [32] without any additional support for sensemaking or reflection. One tool provided a leaderboard that allows students to see how their work ranked compared to classmates [60], but no support for understanding the reasons for their ranking. Even systems that allowed students to enter reflections on the feedback they received found that students rarely used this feature. For example, PeerStudio researchers found that only 100 out of 3600 students wrote reflections using the system [36].

Student/Audience Interaction Systems

To investigate the potential of in-class peer feedback, I reviewed tools designed for collecting input in real time during class. A number of systems support responses to pre-authored multiple-choice questions (e.g. [31]), and some require specific hardware [20].

In-class systems focused on being easy to use and often featured automatic real-time visualizations of student responses. For example, one system allowed students to ask questions during class and used voting as a way to highlight the most popular questions for the instructor [53]. I also reviewed systems designed to poll audiences who were not necessarily students in a classroom. Many systems were similar to in-class response systems in that they supported only multiple-choice responses and visualized responses in real time. Rather than relying on specific hardware like i>Clickers, these tools are often web-based and can be accessed via laptop or mobile devices. For example, Feedbackr, a tool designed to poll audiences during presentations, directs users to a short URL to view multiple choice questions, and the presenter can choose when each question is visible to audience members [19]. These tools commonly solicited anonymous responses from audience members.

Other systems designed to poll audiences went beyond multiple-choice questions. For example, Pol.is is a web-based polling system that uses machine learning and data visualizations to host large-scale discussions [50]. Once a topic has been created, Pol.is allows users to write a response or react (agree or disagree) to the responses of others. Then the system organizes users into groups who share similar opinions. These groups are visualized so that users can see whose thoughts align with their own and which reactions are most important to that group [50].

Lessons Learned from Other Systems

This review of existing systems did not find any tools that focused both on peer feedback and real-time use during class. Existing peer feedback systems require extensive out-of-class time from students and may not promote or allow time for reflection, while existing in-class systems are typically limited to narrow types of interaction. However, both types of tools provide insights for the design of in-class peer feedback systems.

Challenges of Existing Peer Feedback Systems

The current design of existing peer feedback systems limits their effectiveness at engaging students in successful peer feedback for project-based learning. Often the design choices for peer feedback place a burden on students, reduce the likelihood that students will receive feedback in a timely manner, limit the number of perspectives represented in a student's feedback, and reduce the time available to work and reflect on their own assignment.

Instead of motivating students to participate, existing systems burden both feedback providers and feedback receivers. Peer review systems commonly require students to conduct reviews independently outside of class [13, 36] and then perform a revision before a deadline [16]. In the case of PeerStudio, which facilitates rapid peer feedback to students in MOOCs, survey responses indicated that many students felt their schedule was too busy to complete all aspects of this process. Students complained about the workload, saying that "(the instructors) expect us to read some forty page essays, then write the critiques and then review two other people, and then make changes on our work... twice a week" [36]. In addition to burdening students, this reduces the likelihood that students can provide feedback in a timely manner. To scope this, many instructors only require students to review a few peers, which limits the number of diverse perspectives a student will receive in their feedback.

Most existing peer feedback systems do not provide any training for students in how to provide high quality feedback to their peers. In addition, research on existing systems has not found evidence that the quality of student feedback improves over time. This suggests that practicing peer feedback without additional guidance is not enough for students to learn to give helpful feedback.
Most existing peer feedback systems do not help students explicitly reflect on feedback. While some systems permit students to resubmit assignments and get additional rounds of feedback [13], most only allow students to *view* the feedback, without any additional support for sensemaking or reflection. One tool provided a leaderboard where students could see how their work ranked compared to classmates [60], but it did not help students understand the reasons for their ranking. PeerStudio allowed students to enter reflections on the feedback they received, but they found that students rarely used this feature; only 100 out of 3600 students wrote reflections [36].

In summary, existing systems struggle with engaging students in the peer feedback process, preparing students to provide high quality feedback, and encouraging reflection on feedback they received. This work indicates that in-class peer feedback can overcome some of the challenges of conventional peer feedback tools, and provides needed further research on how to better support feedback receivers and providers, as well as instructors.

SOLVING EXISTING CHALLENGES OF PEER FEEDBACK SYSTEMS

Designing Technology for In-Class Engagement

Bringing technology into the classroom may reveal challenges related to classroom culture and technology adoption that are not a factor when using paper to collect feedback. Technology can be a distraction to students during class [22], not only to the students using the system, but also to instructors or students who present to an inattentive audience. Instructors may also be hesitant to introduce an unfamiliar technology into a live class [7].

Although many advocate for the benefits of using paper instead of digital or electronic information (e.g. [57]), researchers have started investigating the effects of using digital systems in the classroom. Piper and Hollan [49] found that students were more engaged with educational material presented on a tabletop display than with the same material presented on paper. The study found students who used the tabletop display solved problems on their own before resorting to answer keys and repeated activities more often than students who used paper documents.
A number of additional factors come into play when considering in-class activities that could impact the design of peer feedback systems.

Time

There is limited time in any class session. Students and instructors may feel overwhelmed with adding another time-consuming activity to the class session [17], so implementing the in-class peer feedback system needs to be simple and time-efficient.

Management Logistics

Setting up and implementing a peer feedback process requires the instructor to manage numerous details, such as assigning peer reviewers or authoring rubrics. The instructor also might also need to evaluate both the original assignment and peer feedback itself in order to provide a grade [39]. (Numerous researchers also explore peer *assessment*, where a student assigns another student a numerical grade on their assignment [e.g. [18]]; however, this paper focuses on qualitative peer feedback, not peer assessment.)

Risks for Students

Peer feedback requires students to make their work public, inviting potential risks such as loss of privacy, saving face, embarrassment, or even humiliation [39]. One indication that students are aware of these risks could be their reluctance to criticize peers when providing feedback [43].

Classroom Culture

Some researchers argue that peer feedback processes can only be carried out effectively when students understand the benefits of peer feedback, trust their peers, and benefit from an established collaborative learning climate [39]. Others point out the necessity for a non-threatening, collaborative atmosphere for peer feedback [48]. The instructor plays an important role in establishing a classroom culture that builds trust among peers and encourages learning without reducing learners' self-efficacy.

Attitudes Toward Technology Use

Most systems for peer feedback rely on digital technology to facilitate interaction, but instructors have a variety of attitudes towards using technology in class. While some instructors find technology beneficial to student learning [45], others decline to use technology in their class because of a lack of technical support or suitable software [7].

Improving the Quality of Peer Feedback

Qualities of Successful Feedback

Regardless of whether feedback comes from instructors, peers, or external reviewers, researchers have identified several key conditions that make feedback more successful for learning.

- <u>Relevant:</u> Feedback should focus on students' learning and performance, rather than on the students themselves or their personal characteristics [24]. Relevant feedback helps students understand the desired criteria (conceptual), compare their actual performance with these criteria (specific), and engage in action that closes this gap (actionable) [55].
- <u>Copious:</u> Researchers and educators value the peer feedback process because it allows students to get a higher quantity of feedback than if the instructor were the only feedback provider [61]. Generating sufficient feedback has most often been limited by the instructors' or peers' time constraints.
- <u>Timely:</u> Feedback should be timely, such that students receive it soon after they submit their work. A recent study on a peer feedback system for large online classes found that peer feedback helped students improve their grades, but only if the feedback was delivered within 24 hours [36].
- <u>Diverse:</u> Seeking diverse feedback helps in many domains, especially in design settings that need to account for multiple stakeholders [4]. Diverse feedback providers are more likely to offer novel perspectives and uncover unique issues [41].
- <u>Reflected On:</u> Reflecting on feedback helps students become better self-regulated learners [8, 39, 46]. Reflecting on feedback requires students to manage their own learning by interpreting standards and rubrics and revising their own work.

Using Rubrics to Guide Peer Feedback

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 [36, 60, 65]. However, even when instructor-authored rubrics are used, feedback providers sometimes still struggle to write specific, critical, and actionable feedback for their peers [36].

Students may learn more from authoring their own rubrics than from using a general rubric authored by the instructor for the entire class [12]. 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 for deeper learning activities according to the ICAP framework [12], as explained below. The potential of rubrics to both increase reflection about what type of feedback is desired and guide feedback providers drives our investigation into how rubrics can best be implemented with in-class peer feedback systems.

Promoting Reflection on Peer Feedback

Reflection encourages students to engage in purposeful thinking, participate in a cycle of inquiry, and form reasoned judgments around a goal [15, 33, 52]. Reflection supports learning in many domains [44], and offers a critical approach to assimilating feedback and improving solutions in project-based STEM learning [56]. In addition, reflection during the peer feedback process can support improvement in students' self-regulated learning skills and help them understand the criteria for success [8, 13, 46].

Researchers have already identified that it is essential for students to reflect so that feedback can be integrated into future work [13, 24, 39]. However, students find reflection difficult, particularly when asked to engage in reflection over time [63]. Existing designs to support reflection include prompts, structured feedback processes, and exposure to high quality reflection [10, 63]. These designs must be sensitively adapted for the peer feedback context, as peer feedback per se does not improve reflection [10], and peer engagement in reflection can sometimes interfere with the reflective process [63]. Current peer feedback systems have not successfully encouraged students to reflect, and researchers have not studied how to encourage integration of feedback into the project-based learning process [36]. We will adapt existing techniques for supporting reflection to a peer feedback context, and study how reflection can most effectively enhance their learning.

INTERACTIVE LEARNING ACTIVITIES

The ICAP framework introduced by Chi et al [12], defines four types of learning activities and argues that interactive learning is the most effective for student learning. Passive learning activities expect students to passively receive information, but nothing more, such as listening to a lecture. Active learning activities encourage students to manipulate the information they receive, such as when a student takes notes during a lecture and organizes the information they receive. Constructive learning activities ask students to not only receive information, but also create or construct something new. For example, after hearing a lecture, a student might draw a concept map of the topics covered, or they might formulate a question to ask about the material. Interactive learning techniques ask students to both *generate* novel learning-related materials and to *collaborate* in engaging with those ideas. For example, two students might debate pros and cons of a topic or discuss similarities and differences. Interactive learning techniques optimize student learning because they allow students to engage deeply in activities such as co-creation [12].

Peer feedback activities at their best can embody the principles of interactive learning, particularly when leveraging the fact that students often work on teams to complete projects. The feedback provider *generates* comments that embody the goals of the assignment and are related to the specific students' work. The feedback receivers then *collaborate* with their team to decide what feedback to implement in the next iteration of their work.

While the generative and collaborative nature of peer feedback produces its strengths as an interactive learning activity, these strengths also introduce challenges that can prevent learners from engaging in interactive activities. For example, because learners are not yet domain experts, the learning-related content they generate may not be accurate or productive. Without some method of feedback on their generated material, students may reinforce their own misconceptions [34]. A second set of challenges has to do with the collaborative aspects of the activity. For example, varying levels of psychological ownership make it more difficult for students to respond meaningfully to each other's contributions as opposed to operating essentially independently [5]. A third challenge relates to the process by which feedback in implemented. Often peer feedback systems place so many time-intensive requirements on students that they feel they do not have the time to collaborative with a team in reflecting on feedback, so the interaction never occurs at all. Additionally, scale becomes a factor. In traditional design crits, students are receiving in-person feedback from a small group of peers, and they are able to have a conversation with their feedback providers during the crit. In typical peer feedback systems when students receiving digital feedback from a larger group of peers, scale prevents feedback receivers from being able to discuss each comment with the feedback provider.

If a peer feedback system was designed to support student engagement, improve feedback quality, and foster interactive reflection, students could get the most benefit out of the peer feedback process. This dissertation describes the process of not only making such a system, but also designing interactive learning activities to support using the system as a way to deepen student learning.

3 SYSTEM DESIGN

This chapter begins by describing the exploratory study where 53 students provided in-class peer feedback using shared Google Docs on student presentations in a project-based innovation course. This study demonstrated that in-class peer feedback could be successful, but that students need better support for reflecting on their feedback. Based on these insights, I led the design and development of a novel web-based, in-class peer feedback tool called PeerPresents. As a preliminary evaluation of this novel tool, fifteen feedback providers used PeerPresents during practice research talks for six PhD students. The preliminary study demonstrated the value of structuring in-class peer feedback, and revealed further insights into how to elicit valuable peer feedback.

This chapter presents the following contributions: 1) an exploratory study of an in-class peer feedback process using an off-the-shelf tool (Google docs), 2) a novel system called PeerPresents designed for in-class peer feedback, 3) a preliminary evaluation of the new feedback tool, and 4) reflections on lessons learned for in-class peer feedback exercises.

EXPLORATORY STUDY OF AN IN-CLASS PEER FEEDBACK PROCESS

The exploratory study of the procedures and design issues for exchanging in-class feedback was conducted by creating a "prototype" from off-the-shelf technology (Google Docs).

Method

I conducted an exploratory study with 53 students (35 female) in a project-based innovation course focused on the design of mobile service applications. The course, offered at a mid-western university, was comprised of undergraduate (37) and graduate (23) students. 95 percent of students regularly carried a laptop with them to class.

Students provided feedback to their peers during mid-term presentations on a group assignment. During two class sessions lasting two hours each, nine groups of students presented their business model ideas for a novel mobile app, such as Friendr, an app to help you find a friend to attend events with you, and SunnySideUp, an app for requesting breakfast delivery to your office or home. To enable peer feedback, each presenting group prepared a business model document that they shared with the class as a Google Doc. During and shortly after each presentation, the instructor asked student feedback providers to write comments using Google's default commenting features, which students accessed using their laptops.

To promote diversity of comments and to scaffold appropriate responses, students were encouraged to comment from one of four framing perspectives:

- Breakdowns: Think of problems that could cause this service to break down.
- Competitors: Think of existing and potential competitors to this service.
- Stakeholders: Think through the perspective of the people involved: users, providers, investors, marketers, and local businesses.
- Scenarios: Think of additional scenarios where this idea could be applied.

These four framing perspectives were intended to stretch students to think about different factors that could affect the design of a mobile app. Students were assigned a different perspective for each presentation.

I collected and analyzed student and faculty comments on each presentation. After both sessions ended, students filled out an anonymous online survey about the experience; 84% participated. I collected data about class participation, students' technology use during class, and their attitudes about the perspectives they were assigned. Students were also asked to compare their participation in providing feedback during the presentations with how they provide feedback in other courses. When surveyed about Facebook use during a typical class, 40% of students reported using Facebook in class once a week and 27% reported using Facebook every class session. 24% reported once a month use, and only 9% said they never use Facebook in class.

Results

In presenting the results, I combined analysis of comment data with survey results to highlight five important criteria for student feedback that emerged from the literature review: relevant, copious, timely, diverse, and reflected-on.

Relevant

80% of students who responded to the survey said that the comments they received were "helpful" or "very helpful." Students' beliefs about the helpfulness of comments were a useful proxy for relevance; they are the ones who must interpret the comments and decide how to use feedback to revise their work.

Copious

Overall, 36 out of 53 students (68%) commented at least once during the two class sessions. Across the nine presentation documents, students made 242 comments, for an average of 26.7 comments per document and 7 comments per student. Of the 36 students who commented at least once, 33 (90%) reported that they felt they gave more feedback using written comments than they would have given verbally in class.

Timely

88% of feedback was given during class, and was available to the presenting group immediately after their presentation. Of the feedback given after class, most was from the instructor giving grades on the assignment. Twelve student comments arrived after class had ended.

Diverse

77% of comments used one of the four framing perspectives. However, only 14% used the perspective assigned to the student at that time; 63% of student comments came from a different perspective than the one assigned, and the remaining 23% of comments were not from any of the perspectives. In other words, students commented from a variety of perspectives even when asked to focus only on one. 38% of students reported their assigned perspective was "difficult" to adopt, and only 40% found the perspective "helpful" for inspiring new ideas while commenting. 77% of comments came from one of the assigned perspectives – even when not assigned to that student. Students found the scaffolding helpful, but felt restrained by commenting from one perspective.

Reflected On

Out of 242 total comments, presenting students made only 5 comments in reply to the feedback their team had received. The data collected does not provide enough information to draw conclusions about how students reflected on peer feedback.

Discussion

The exploratory study provides preliminary support that in-class feedback can be relevant, copious, timely, and diverse. This was achieved without implementing review requirements, mandating additional time outside of class, or requiring specific hardware. However, there is less evidence regarding how students reflected on the feedback they received. Google Docs collapses longer comment threads, making it difficult for students to read all feedback on a document, or even know how much feedback is available. Google Docs also did not explicitly support reflection processes, and so future systems should address this need.

While in-class feedback did result in a diverse set of student comments, students did not seem to use the prescribed perspectives. Students were instructed to comment from a single perspective, but instead commented from many perspectives – creating even more diversity of feedback than we tried to scaffold. Students reported they did not use the framing perspectives as prescribed, partly because it felt limiting. I hypothesized that students are both willing and able to adopt multiple perspectives while commenting on a presentation. However, scaffolding may still be necessary to ensure that providing diverse feedback perspectives remains relevant to presenters' needs and the instructor's goals.

Despite initial concerns about placing a cognitive burden on students in the evaluation study, preliminary observations revealed that students have a cognitive surplus during class, especially during peer presentations, which they often spend on non-academic activities like checking Facebook. A well-designed in-class peer feedback system could give students a pedagogically meaningful activity to potentially replace Facebook and other distractions with academic exercises. Finally, the prototype did not support all students equally effectively. Most students felt comfortable giving and receiving comments on their work, but a minority of students (7 out of 45 survey respondents, or approximately 15%) felt that commenting during the presentation was distracting for the presenter. Additionally, some students reported feeling overly criticized and judged, or that commenting felt like a "hazing" process. This feeling of being judged may make peer feedback less effective and reduce learners' willingness to participate. Future systems could do more to address student risks and ensure peer feedback remains focused on the work, not on the personal characteristics of the student.

To address the findings from the exploratory study, I expanded the notion of in-class feedback to encompass the process before, during, and after class.

DESIGN OF A NOVEL IN-CLASS PEER FEEDBACK TOOL

The exploratory study validated the concept of exchanging peer feedback during class using web technology, and I gained insights on how to design a system that supports relevant, copious, timely, diverse, and reflected-on feedback. In creating PeerPresents, I focused on features that would build on the successes of the exploratory study while reducing the burden on the instructor, providing

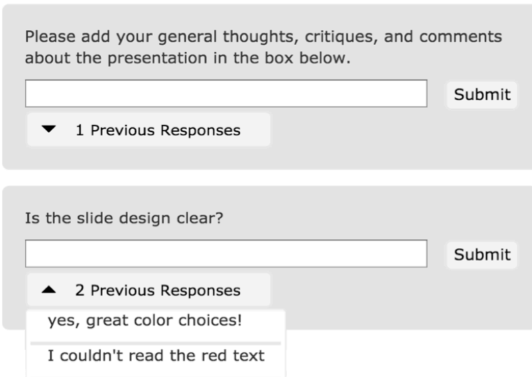


Figure 1. Peer feedback interface: Peers respond to specific questions as many times as they desire.

Figure 2. Students agree or disagree with the responses of other students providing feedback.

scaffolding to encourage relevant comments, mitigating risks for students, and encouraging reflection on feedback.

Before Class

Question scaffolding

To create effective scaffolds without placing an additional burden on the instructor, PeerPresents allows each team of presenting students to author guiding questions in preparation for their presentation. These questions are presented to the class during the presentation, allowing presenting students to receive feedback better tailored to their goals. Not only do the questions provide a guide for students giving feedback, but the process of generating questions also encourages teams to reflect on what type of feedback would be most useful, and how to ask effective feedback questions. Additionally, structuring feedback around questions rather than around, for example, slide numbers allows our system to accommodate presentations of many types – formal slide-based presentations, in-class role-play, video pitches, and more.

Permissions and Dissemination

In PeerPresents, students can add project group members to their presentation, so that the instructor does not have to input teams for the entire class. The system also automatically creates a short URL so presenters can easily direct feedback providers to their presentation.

During Class

Pseudonyms

To help reduce the perceived risk around peer feedback, while still providing a degree of accountability, PeerPresents allows feedback providers to choose a pseudonym when accessing the system. This pseudonym appears on every comment they make, which is visible both to other students in class for voting and to the presenters for reflection. Because students can choose obscure or identifiable pseudonyms, students can decide in each class session how anonymous they would like to be.

Voluntary Feedback and Voting

During the presentation, peers can answer each question provided by the presenters as many or as few times as they choose (see Figure 1). They can also choose to provide comments in a "default" open-ended text box. PeerPresents did not support discussion threads between students to minimize distractions during the presentation.

Peers can up-vote or down-vote comments from other students; this helps the presenters identify comments that are important, popular, or controversial. Comments are displayed in real time throughout the presentation on a separate voting page (see Figure 2).

After Class

Reflection

Presenting students can immediately view all their feedback with the timestamp, author, up/down votes, and framing question (see Figure 3). Students can then tag and filter comments to organize their feedback in a meaningful way. The system provides default tags for students to mark Positive, Negative, and Important feedback. Any presenting team member can individually create additional tags, which are visible to all team members. The system also displays how many comments have been recorded under each author, question, or tag. Students can filter comments by author, question or tag, and sort comments by least or most recent, author, and number of votes. To increase familiarity, we designed the filtering and sorting process to resemble many consumer product websites.

Not only are these features novel within the design space of peer feedback systems (see Table 3, above), they also work together to address the issues raised by the exploratory study and to improve the peer feedback process.

Figure 3. Presenting students see all the feedback they received, with the ability to tag, sort, and organize comments. (All names and pseudonyms removed for anonymous submission)

PRELIMINARY SYSTEM EVALUATION

To evaluate the system, I asked six PhD students to use PeerPresents while practicing for a PhD requirement talk—a twenty-minute research presentation followed by a ten-minute question period. This presentation is a requirement for the doctoral program in their academic department. Faculty members judge each presentation on research content, communication style, and slide design.

Methods

Six students received comments on their presentations, and fifteen people (13 students and 2 faculty) provided feedback. I collected survey data asking participants to comment on their experience with the system with respect to the relevance, quality, timeliness, diversity, and reflectiveness of feedback, and I analyzed the feedback provided by peers and faculty. The research team developed metrics to categorize comments into the following five types of feedback: *On Topic*: 1) research content, 2) communication style, 3) slide design, and *Off Topic*: 4) emotional support, and 5) other. I also tracked whether feedback made reference to a particular slide by number.

The practice session happened five days before the final presentations, so the talks were still in progress at this stage. This also meant that comments on communication style and slide design would be highly relevant, even if they might not be as important for practice talks in other contexts. Given the constraints of the presentation context, I made some modifications to PeerPresents. First, the PhD program has a specific form for providing written feedback on these talks. Rather than have presenters develop their own questions to scaffold feedback, I used the same questions faculty would use to evaluate the real talk. For example, students would be evaluated on how the research was "situated in a larger theoretical context", if the research was "communicated in an understandable way", and whether "the slides followed good aesthetic design principles." All students would be evaluated by the same rubric.

After the practice session, I asked presenters to reflect on the feedback in a spreadsheet and to use the filtering features to mimic tagging and sorting their feedback. (This informed key features for the PeerPresents reflection page, which was under development at the time.)

Results

Relevant

Five out of six presenters responded that the feedback was relevant. In coding the comments across all presentations, we found that 89% of comments were on topic (45% content, 15% communication, 29% design) and only 21% were off topic. (Total percentages sum to over 100% because some comments included multiple types of feedback.) Despite concerns that off-topic feedback might be negative or judgmental, we found that the vast majority of off-topic feedback—90% of all off-topic comments—provided positive emotional support. Only 2% of comments were completely unrelated to the presentation or presenter (e.g. "hi" or "Lightning bolt! Lightning bolt!"), and none were negative.

I also found that 33% of the comments mentioned a specific slide number. Mentioning specific slides seemed particularly salient to presenters. Students commented that they "appreciated the feedback on specific slides" and "got specific details that I would not have gotten from discussion."

Finally, I found that instead of answering all questions on the feedback form, students put 44% of their feedback into first question visible at the top of the form, asking about the content of the talk in general. The second most frequently answered question (asking for feedback on communication skills) was the form's second question; it received 13% of all feedback.

Copious

Across all six presentations, there were 338 comments. Remarkably, feedback providers gave an average of 24 comments, something that would not have been possible during a 10-minute verbal feedback session. In the open-ended survey, participants responded that they "got a lot more feedback in this form, than traditional ways" and that it made it "easy to give lots of feedback."

Timely

All commenting happened either during the 20 minute talk or within the 10 minute Q&A period immediately following each talk. The total session for all six talks lasted three hours. According to our debrief discussions with presenters, all participants were eager to see their feedback immediately after their talk. When students knew that feedback could be available immediately after their presentation, they preferred not to delay even half an hour while another student spoke.

To minimize distraction, I asked students providing feedback to answer questions throughout the twenty-minute research presentation but wait to vote on others' comments until the ten-minute Q&A period. Participants expressed a desire to see others' responses to questions sooner, rather than waiting until after the presentation had ended.

Diverse

Participants were pleased with the diversity of their feedback. In addition to receiving comments regarding presentation skills, slide design, and research content, presenters also commented that they received a mix of detail-oriented and general feedback, and a balance of comments stating a problem and comments offering a solution. Presenters also recognized that the diversity in their feedback was in part due to the number of diverse perspectives represented in the audience.

Reflected On

I asked presenters to use the filtering features in a spreadsheet to mimic the feedback tagging and sorting features in PeerPresents. While not all presenters used the tagging feature, all mentioned frequently referring to their feedback during revision. One presenter told researchers her feedback contained a lot of "why" questions asking her to explain the validity or reasoning behind specific statements she made in the presentation, saying, "[the feedback] tells me I haven't found the right words or the right slides to tell the story yet."

Feedback providers voted on 120 of the 338 comments; 177 total up-votes and 7 total down-votes were given. The average feedback provider voted on 12 comments. Of the 113 comments receiving up-votes, 73 (65%) received only 1 vote and 30 (27%) received only 2 votes. No comment received more than one down-vote.

Most participants entered their real name as their pseudonym in the system. Twelve out of fifteen participants used their real name, or a variant on it, including five out of six presenters.

DISCUSSION

The preliminary evaluation of PeerPresents demonstrates that a custom in-class peer feedback system can yield relevant, copious, timely, and diverse feedback, as well as provide time and tools for reflection. In addition, this system addressed student concerns about negative personal judgments and minimized the burden on the instructor in setting up our system. I also identified a number of issues for future research and development.

Peer Feedback as Legitimate Peripheral Participation

As described above, I observed copious feedback being provided to presenters by their peers. However, I also observed that our system supports a variety of different levels of engagement and participation. Three feedback providers gave significantly more feedback than the average, with 80, 71, and 66 comments respectively. Others made only a few comments; four feedback providers made fewer than five comments apiece. The type of comments also varied: all fifteen gave feedback on slide design, nine spoke about communication skills, twelve about research content, and thirteen gave emotional support.

Building on theories of legitimate peripheral participation, this diversity of participation styles can be considered a measure of the success of our system. Legitimate peripheral participation describes how novices become experts by participating in simple tasks that are well within their capabilities, but as part of a larger community in which they can receive feedback from more experienced members [37]. Similarly, PeerPresents does not require individual feedback providers to give a certain amount of feedback, but rather lets individuals choose the amount and type of feedback they are comfortable providing. This theoretical approach, however, suggests that it is important to make expert behavior visible to novices for modeling and learning purposes. For example, we can investigate the impact of design decisions such as highlighting faculty feedback with a different color.

Challenges of Feedback Questions

The PeerPresents system uses questions to inform peers about the feedback most relevant to presenters, and to remind them about presenter needs as they give feedback. The system succeeded at focusing the feedback providers, as the vast majority of feedback was relevant. However, more than half of it was given in response to the first two questions on the feedback form. These two questions received a variety of on- and off-topic feedback. The first question was a place for general open-ended feedback, and about 55% touched on the basic research content (as opposed to presentation style or slide design). The second question asked for input on communication skills, and here only about 59% of feedback touched on the presenter's communication skills and slide design. This suggests that feedback providers were able to use questions to scaffold relevant feedback, but that they sometimes provided feedback that did not match the question being asked (e.g. 36% of responses to the second question were about research content rather than communication skills). Providers also chose not to respond to questions later on the list.

I infer that feedback providers behaved this way because the feedback form contained nineteen questions. With so many questions, finding a particular question meant scrolling through multiple pages – time that feedback providers preferred to spend commenting. This is in line with prior work that suggests students often do not read rubrics in their entirety [2]. Additionally, the questions were intended to evaluate a finished presentation, not a work in progress. Students remarked that while there were too many questions, they also did not cover everything they wanted to comment on. Searching for a question that might not exist may have seemed like a fool's game when an open-ended comment box was convenient.

I found that the system worked, but that the question scaffolding may need further consideration. There may be a practical upper limit on how many questions commenters can effectively respond to. I did not test what makes an effective question for feedback, but existing research on rubric formation (e.g. [3]) could provide guidance for helping students construct questions to guide feedback in future iterations of the tool (see Chapter 5).

Limitations on Voting

On average, feedback providers made twice as many comments as votes, even though commenting required more time and effort. General comments (like "you may want to talk slower"), or comments that were easy to agree with (like "you are a great speaker"), received the majority of the up-votes. Perhaps this happened because many of the feedback comments lacked context after the fact. For example, some feedback providers said they could not agree or disagree with someone else's comment because they did not remember the slide being referenced. Even when a comment did not mention a specific slide number, there was not always enough context after the fact for feedback providers to agree or disagree. While presenters were generally able to make sense of comments by referencing their own slide deck or recalling their presentation, the other feedback providers lacked the context they needed to be helpful up voters. While this difficulty may have been exacerbated by using the departmental feedback form instead of presenter-authored questions, the reconstruction of context is a future challenge for this work.

Very few of the votes were down-votes – only seven out of 184 votes. The lack of context may have made feedback providers even more hesitant to criticize than to applaud other people's comments, since they might be lacking context that would make the comment productive. Other research suggests that providing only upvoting, instead of upvoting and downvoting, can help minimize evaluation anxiety for students in a peer feedback context [30]. In this case, removing downvotes would be no loss, since students did not use them, but could be a gain for classroom culture. Removing the option to disagree with students in PeerPresents might benefit the classroom culture by removing the potential fear of getting a downvote and helping students feel more comfortable with the peer feedback process.

We also note that 35% of comments that pointed out problems also included a suggestion for how to address it; this could help mitigate the criticism. Taken together, this data makes us ask whether students are being sufficiently critical of one another's work, or whether additional scaffolding is needed to help feedback providers express appropriate criticism of the comments provided.

Pseudonymity and Learning Culture

While participants had the option to use pseudonyms as identifiers in the system, the majority of participants chose to use their real names, or variants on their real names. While being identifiable may produce social risks for students in some contexts [11], in this study, being identifiable was seen as a positive. Because they personally knew each feedback provider, presenters found their knowledge of that person's background and expertise helped provide context for specific comments. Having the option to be identifiable also served as a partial way of allowing experts to model behavior. Because the faculty members attending the talk chose to be identifiable, students were able to give their comments higher priority.

This does not imply that identifiability is necessarily better that pseudonymity. This study observed a particular learning culture, one in which all participants know each other well and had established trust before the study. This type of learning environment is conducive to productive peer feedback [39], and future work could investigate how PeerPresents can foster such a classroom culture in a variety of learning environments, while still supporting learning environments that do not yet have these characteristics.

Class Size

While PeerPresents may be particularly helpful to instructors of large classes, our small-scale preliminary evaluation indicates that high-quality peer feedback can be useful for classes of all sizes. The 15 students in the preliminary tool evaluation were able to instantly give a detailed critique of their peers' work, allowing presenters to rapidly iterate and improve their presentations.

Distraction

One concern with using this tool in real-time is that the system could become a distraction during student presentations. For example, students might be reading and voting on peer's comments and stop listening to the presenter. Also, how will the presenter feel if everyone in the audience is looking at their laptops instead of at the slides or speaker? In the preliminary study of the system, participants did not express distraction as a concern, neither when they were giving feedback nor when they were presenting. Perhaps this is because students in our study are accustomed to being on laptops during presentations. However, this might not be true for all classroom contexts. A fruitful avenue for future work will be to investigate how facilitate minimally disruptive in-class feedback.

Role of the Instructor

PeerPresents does not specify what role the instructor plays while students are using the system. The instructor could use PeerPresents to help students develop good feedback skills, such as guiding students as they write feedback questions, providing comments through the system during presentations, and evaluating the quality of the feedback students provide. The system might also provide data visualizations to the instructor to provide a clearer picture of how students are using the system and what additional support they might need. Further design work can explore the best ways to support how instructors might interact with PeerPresents.

Limitations

This study recruited a limited sample to test PeerPresents, consisting of graduate students and faculty from a single department. The tool was implicitly compared to a verbal Q&A session, rather than testing it in comparison to a similar system, or to other existing approaches such as email or paper-based feedback forms. Further, PeerPresents' reflection page was still under development during the initial evaluation; participants may have organized their feedback differently in Excel than they would when using the PeerPresents reflection page. In addition, the exploratory study and preliminary system evaluation did not fully investigate the instructors' perspective towards implementing this type of system in their class. I expect to gain insights about what instructors need from this system when we deploy PeerPresents in classrooms.

SUMMARY

This chapter introduced a novel system for in-class peer review called PeerPresents. Through an exploratory study with Google docs and a preliminary evaluation of PeerPresents, I demonstrated that in-class peer feedback systems can elicit relevant, copious, timely, and diverse feedback through a structured in-class activity. By allowing students to generate scaffolding questions before class and enabling feedback provision during class time, the system maximizes the time for student reflection on feedback after class. Students also described the feedback they received as helpful and reported that they gave more feedback than without using the system. These early results demonstrated the potential benefits of in-class peer feedback systems, highlighted a new design space for classroom tools, and extracted generalizable design lessons for future work.

4 STUDY 1 – ENGAGING IN PEER FEEDBACK

After designing the PeerPresents system, I next set out to validate the system and ensure in-class peer feedback is a viable process for student learning. Because in-class peer feedback systems are a novel technology, prior work has not sufficiently explored the potential drawbacks of digital in-class peer feedback. For example, bringing new technologies into a classroom might be distracting for students and unwieldy for instructors.

Instructors lack guidance on the efficacy of using devices to facilitate in-class peer feedback and how this compares to other real-time feedback methods. Therefore, this study investigates if a digital peer feedback system produces feedback of equal or better quality and quantity compared to a paper-based method during class, if the digital system distracts students, and if the digital system provides additional benefits. When analyzed through the lens of ICAP theory, we would expect students to provide equal feedback quality and quantity, since writing on paper and typing in a digital system are both considered constructive learning activities. While we are validating one particular system, our findings could be generalized to other contexts where student teams present unfinished projects in class and receive both instructor and peer feedback on their work.

I conducted a within-subjects study in a project-based course with 73 students. During two consecutive weeks of project presentations, peers provided feedback either on paper or through a digital device. To facilitate digital feedback exchange, we used PeerPresents. Students provided feedback (either on paper or a device) while the instructors gave verbal feedback, keeping with the culture of this particular course.

This study examined the quantity and quality of feedback provided with each method, student perceptions of each method, and differences between student and instructor feedback. We found that both methods yielded a similar number of unique open-ended comments, although the digital approach led to more student participation overall when accounting for upvotes. Students generally felt they had better feedback exchanges on paper, although assessments on the feedback by instructors indicated no difference in quality. Surveys and interviews with students revealed split opinions. Students who preferred paper liked the ability to create sketches; others appreciated the social awareness and dynamics afforded by the digital method.

This chapter contributes: 1) empirical evidence that digital systems can be equally as effective as paper-based methods for eliciting student feedback during class, 2) design implications for peer feedback systems, and 3) a discussion on how classroom culture and student expectations might affect the implementation and use of an in-class peer feedback system.

METHOD

Participants

We conducted a study with 73 students in a graduate-level project-based course on designing virtual environments and games (see Figure 4 for an example). The instructor divided students into fifteen teams to complete course projects, each lasting around two weeks. Teams were randomized between projects. The course was divided into two sections—one in the morning and one in the afternoon.

Figure 4. Screenshot from a game created by students in the final presentation.

Students were assigned to present their work in a particular session, but they could also attend the other session to provide feedback, and often did. Each section met for three hours twice a week. All fifteen teams presented designs to the class each week, either as a work in progress or as a finished product. Half of the teams presented in the morning section and half in the afternoon section; in both sections, the entire class session was devoted to presenting work and receiving feedback.

Classroom Setting

Before the comparative study, I observed the class for one week to establish a baseline before introducing the digital system. I also briefly interviewed the two course instructors prior to our observation using a semi-structured interview process to gain insights on their development of the existing peer feedback techniques used in this course.

As part of the regular feedback process in the course, the class used a paper-based system. Before class, the teaching assistants would leave a pile of notecards on each chair in the classroom. Students were expected to write feedback on one notecard per presenting team. Figure 6 shows an example paper notecard. After each student presentation, the course instructors would give public verbal feedback to the presenting team. Students could provide written feedback at any point; there were no restrictions on when they could and could not write. At the end of class, notecards were handed directly to a representative from each presenting team. Instructors and teaching assistants never saw what was written on the notecards, and did not record or grade participation in the feedback process.

Despite the existence of a formal feedback process, the course instructors did not feel students were successfully learning peer feedback skills. Instructors commented that students would not give helpful or high-quality feedback to each other. Further, the instructors felt the teams were not sufficiently reflecting on the instructors' or peers' feedback. One instructor pointed out that they had recently implemented a new assignment to "make them at least read [the feedback]".
An instructor said that "the class is defined by students showing work off to each other" and that they believe students have a "strong desire to share and discuss what's being done." I observed students paying close attention during student presentations, so much so that they would often audibly react to the game demos. After each presentation, instructors would spend up to twenty minutes giving verbal feedback. During this time, students would spend no more than five minutes writing feedback on their notecards, and I observed how students would drift towards Facebook, mobile phones, or other off-topic distractions until the next presentation began. I frequently observed students falling asleep between presentations while instructors were speaking.

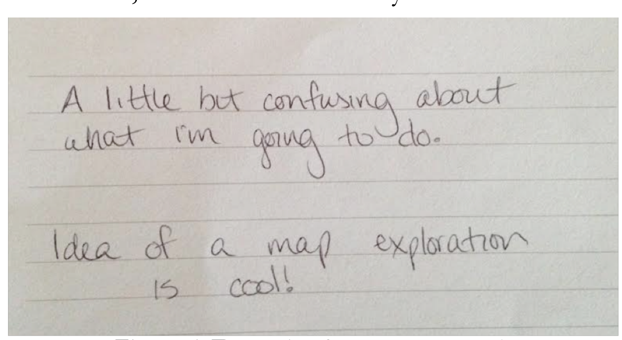


Figure 6. Example of a paper notecard.

Procedure

We implemented a two-week study during the final project. During the first week, the morning section gave feedback on paper, and the afternoon section used the PeerPresents system to provide feedback. For the second week, the afternoon section gave feedback on paper, and the morning section used the digital system. Each team presented for about four minutes, followed by 10-20 minutes of instructors' verbal feedback, during which time students provided feedback on paper or through their devices. For these two weeks, if a class section was in the digital condition, the teaching assistant for the course transcribed all the instructors' verbal comments into the digital system for later comparison with student comments. The research team had extra devices available in case students did not have access to technology, but the students never needed them.
Students were given a pre-survey before the study asking their opinions on the paper process, challenges they faced when giving or receiving feedback on paper, and their current process of reflecting on feedback. The pre-survey consisted primarily of Likert-scale questions for agreement: for example, "How much do you agree/disagree with the following statement: I can say everything I want to say using the paper notecards". Students were also given a post survey at the end of the study asking them to compare the paper process to the digital system by reporting perceived differences between each method with multiple-choice questions (PeerPresents, Notecard Rodeo, or No preference). We also asked students to express their preferences on open-ended questions, which we analyzed to extract themes and interesting comments.

Measures and Analysis

Student Participation

From system logs, we collected the usernames used by students, the responses and timestamps of those responses, and a list of usernames that voted on each response. We collected and transcribed the notecards and counted how many notecards included drawings.

To measure how long students stayed engaged with paper-based peer feedback, a researcher observed two class sessions before the digital system was introduced. The researcher sat such that she could unobtrusively observe all students. The researcher recorded the times when each presentation started, when the majority of students had stopped writing, when all students had stopped writing, and when the instructors stopped talking. Length of engagement with the digital system was computed using system logs.

Feedback Quality

We surveyed students on the perceived quality of the feedback exchange with each method. We also asked students for their opinions on the strengths and weaknesses of each method. Finally, we randomly selected 10 comments each from two digital condition presentations and two paper condition presentations and asked instructors to rate the quality of comments on a scale from 1 to 5. Instructors were blind to condition; the comments were a random sample of 10-15% of comments from 25% of the teams, as instructors were under tight time constraints and could not rate all comments.

Comment Relevance	On Topic: Content	"Music could have been more intense"
	On Topic: Communication	"It's interesting how it's only one person responding to questions, and the same person standing apart from the
	Emotional Support	"Great effort"; "I would love to try your game"
	Off Topic	"Shooting lazers!!!"
Comment Form	Descriptive	"health bar looks like a shield"
	Actionable	"Could add sound effect for teacher turning around"
	Question	"What was the punching/bosing glove? How was it to
	None of the Above	"RIP mama bird"
Comment Sentiment	Praise	"Whoa nice lighting effects"
	Criticism	"The animation of policeman doesn't stop at right
	Neither	"How to lose the game?"

Table 4: Example comments for each coding category.

Content Analysis

To further investigate the content of written responses, we segmented responses into individual comments and coded them into categories (see Table 4). We divided responses into individual comments primarily based on physical markers such as bullet points or numbers, separate sentences or sentence clauses, and skipped lines between items. We coded feedback into categories based on relevance, form, and sentiment (see [46, 55]).

For relevance, comments were coded as On Topic (focused on content), On Topic (focused on communication skills), Emotional Support, or Off Topic. Comments were also coded as either Praise, Criticism, Neither praise nor criticism, or Both praise and criticism. Comments were also coded as Descriptive, Actionable, Question, or None of the above. Descriptive comments were stating observations or describing the game. Actionable comments were offering specific suggestions or giving instructions. Questions were comments where students asked the presenters a question. Table 2 lists examples of each type of comment. Three independent coders categorized 10% of the data and achieved moderate agreement. After discussion and retraining on all categories, coding agreement on an additional 5% of the data reached an acceptable agreement of ICC 0.85.

RESULTS

Student Behavior

The overall results show that students wrote a similar number of comments on paper as the digital condition, and instructors rated comments from both conditions to be equal in quality. Students in the digital condition interacted in more diverse ways and for a longer period of time than with paper, despite students expressing a preference for paper.

In both the paper and digital condition, there was an average of 20 respondents per presentation (approximately 2/3 of the class; the other 1/3 chose not to participate). Each respondent gave an average of three comments per presentation (an average of 57 comments per presentation). About half (56%) of students' digital responses received at least 1 upvote. Across all presentations in the

		Student notecards	Student digital	Faculty
Comment Subject	On Topic: Content	793 (92.1%)	666 (82.5%)	149 (98%)
	On Topic: Communication	20 (2.3%)	35 (4.4%)	2 (1.3%)
	Emotional Support	43 (5.0%)	33 (4.1%)	1 (0.1%)
	Off Topic	5 (0.5%)	64 (8.0%)	0 (0%)
Comment Form	Descriptive	642 (74.6%)	535 (67.0%)	96 (63.1%)
	Actionable	113 (13.1%)	109 (13.7%)	44 (29.0%)
	Question	75 (8.7%)	91 (11.4%)	12 (7.9%)
	None of the Above	31 (3.6%)	63 (7.9%)	0 (0%)
Comment Tone	Praise	406 (47.2%)	311 (39.0%)	30 (19.7%)
	Criticism	227 (26.4%)	200 (25.1%)	53 (34.9%)
	Neither	269 (31.2%)	318 (39.8%)	75 (49.3%)

Table 5: Distribution of comment type for students in each condition and for instructors.

digital system, 44% of students were active voters (more than 3 upvotes), and 1 student voted on every presentation but did not provide any written comments. 13% of paper notecards had drawings. Student's drawings were not substanitive and they did not serve as a way to provide additional feedback. Instead, drawings were more like stickers; they were viewed as a way to congratulate or motivate fellow classmates and friends in a fun way.

No Difference in Number of Comments per Condition

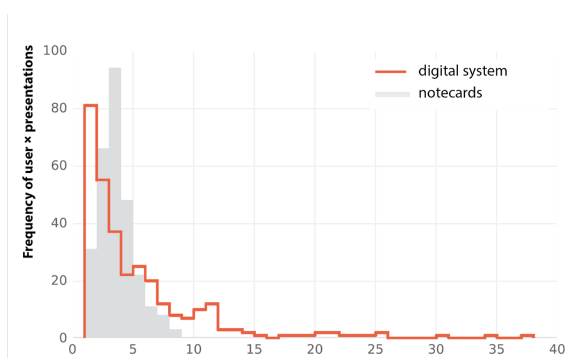
Students produced a similar number of textual comments on paper as they did on their devices (for paper, M=61.5 (SD=18.75); for digital, M=53.2 (SD=17.42)), and there was no significant difference between conditions in the average length of comments (for paper, M=50.07 (SD=38.92); for digital, M=55.94 (SD=48.45)). There was also no significant difference in the number of comments per presentation (for paper, M=3.04 (SD=1.45); for digital, M=2.86 (SD=1.85)). Based on our observations and from our discussions with students, students did not give verbal feedback to other teams at any time outside of class.

No Difference in Quality of Comments per Condition

Based on instructors' blind-to-condition ratings on a scale of 1 to 5, there was no significant difference in comment quality between conditions (for paper, M=2.7 (SD=0.77); for digital, M=2.5 (SD=0.65)).

More Ways to Interact in Digital Condition

The digital system provided an additional way for students to interact through upvotes. Figure 7 shows the distribution of interactions per user over time. With the digital system, many students made only one interaction, while one made as many as 37 interactions in one presentation.



Interactions per user per presentation
Figure 7. Distribution of interactions per user per presentation shows some students interacted much more with the digital system.

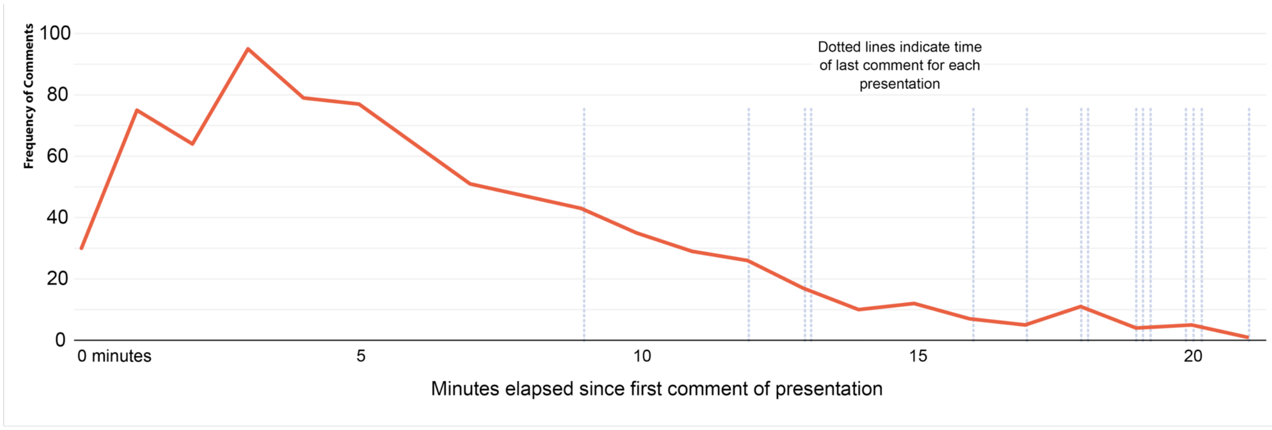


Figure 8. Digital comments peaked around 4 minutes, but continued for up to 20 minutes for some presentations.

Longer Length of Engagement in Digital Condition

We compare the time when the last student stopped writing on paper to the time when the last student stopped commenting in the digital system. In our observations of the paper notecards, the majority of students began writing comments after the presentation ended and stopped participating in the feedback process within five minutes (M= 3.75 minutes (SD=0.96)). We observed that all students had stopped writing on notecards after 8 minutes. With the digital system, students stayed engaged longer and spent more time giving comments, as shown in Figure 8, some continuing to comment for more than 20 minutes.

Students Thought Paper Produced More Feedback

On the post survey, students reported feeling that they received more feedback from paper notecards, $X^2(1,N=73)=17.31$, $p<0.05$. Students did not report a significant difference between paper and digital in terms of what led them to participate more during class, or what led them to give more feedback during class.

Students Perceived Paper Produced Better Feedback

Students believed they gave higher quality feedback on paper, $t(73)=8.33$, $p < 0.05$ Students also believed they received higher quality feedback from paper notecards, $t(73) = 35.0$, $p < 0.05$. However, as noted above, there was no significant difference in comment quality between conditions.

More Content-Related Comments Using Paper

There was a significantly different distribution of comments among relevance categories, X^2 $(3,N=1659) = 55.60$, $p<0.05$ (see Table 5). Students in the paper condition had more On-Topic-Content comments, fewer On-Topic-Communication comments, and fewer Off-Topic comments than students in the digital condition. There were a similar number of Emotional-Support comments, Critical comments, and Actionable comments between conditions.

Faculty Gave More Critical and Actionable Feedback

Instructors gave more relevant comments, less praise, more criticism, and more actionable comments compared to students in either condition (see Table 5). This aligns closely with the prior work's definition of "good" feedback as relevant, critical, and actionable [46, 55].

Student Opinions

Open-ended questions on the pre and post surveys asked students to describe the strengths and weaknesses of each feedback method. Students provided opinions on four key aspects.

Physical Constraints on Paper and Devices

Students said that it was hard to read handwriting on the notecards, and we observed students ignoring cards they could not read in team discussions. Students said it was difficult to write in the dark classroom, and they frequently forgot to bring a pen. Students also believed they ran out of space on the cards.

When using the digital system, students said that not everyone has a laptop, and the mobile view of the system was not always sufficient. Students expressed the inability to draw as a major shortcoming of the digital system.

Digital Supported a Social Atmosphere

Students felt that the digital system was "fun", "funny", and "social." Students commented that they liked being able to see other students' opinions during class. However, some students felt the social atmosphere had drawbacks. Students felt the comments turned into more of a group chat than a meaningful critique, and that some students were "performing" by trying to write the funniest or most provocative comments, rather than trying to give the most helpful feedback.

Anonymity and Criticism

A few students commented that they felt uncomfortable being critical when the entire class could see their comments, even when the comment did not have their name on it. Students also disagreed on which method felt more anonymous. Some felt notecards were "completely anonymous", even though we observed that students often recognized each other's handwriting and drawing style. Others appreciated that the digital system allowed them to choose an anonymous username, and that the username was only visible to the presenting team.

Organization in Both Conditions

There were dissenting opinions among students about which method better supported feedback organization after class. Some felt notecards were easier to organize because all team members could see them without needing a device. Others preferred the sorting and tagging features provided in the digital system. Others did not organize their feedback with either method.

DISCUSSION

This within-subjects study compared different media—paper and digital—for collecting peer feedback in a classroom setting where student teams present unfinished projects. Students wrote a similar number of comments with both methods, and instructors rated comments from both conditions to be equal in quality. As predicted by ICAP theory, both paper and digital activities are constructive learning activities and should produce similar levels of learning.

Despite instructor assumptions that technology in class would be distracting, we show that in-class technology that supported course learning goals was not distracting to students. In fact, students in the digital condition interacted in more diverse ways and for a longer period of time than with paper.

However, students generally preferred the paper notecards, and believed that they received more feedback and gave higher quality feedback on paper than online. Here we discuss the discrepancy between students' perceptions and reality, the ancillary benefits of going digital, the effects on classroom culture, and the pedagogical affordances of each approach. Finally, we deliberate on the design implications, study limitations, and future work.

Student Perceptions vs. Reality

Students accurately believed that they *gave* the same amount of feedback with each approach. However, they inaccurately believed they *received* more feedback from paper notecards, even though there was no significant difference in the average numbers of comments per presentation between conditions or the average length of comments. Our observations revealed that some students did not bring a device to the team meetings when students discussed their feedback, which may account for this discrepancy. Even though an equal amount of feedback was available, some students may not have seen it.

Students also believed they gave and received higher quality feedback on paper, and that the digital system was more of a social group chat than a meaningful critique. However, the instructors' ratings indicate no difference in comment quality between conditions. This misconception may be due to the different distribution of comments for each method, as students in the paper condition gave more content-related comments, fewer communication-related comments, and fewer off-topic comments than in the digital condition. This misconception could relate to the students' perspective that peers were acting like "performers" on the digital system. The public nature of commenting may have led students to believe they received more off-topic comments than was actually the case.

Ancillary Benefits of the Digital System

Paper notecards and the digital system perform equally well at generating useful comments. However, there are additional advantages to using a digital approach.

Facilitating Diverse Input and Social Interactions

The ability to comment quickly, endlessly, and in real-time seems to have encouraged students to give feedback on in-the-moment observations such as presentation style, which they were less inclined to comment on when using paper, perhaps because of their perception of a limited response space. The digital system allowed students to comment anonymously if they chose, which students valued.

Importantly, the digital system allowed multiple types of interactions during class, a strategy supported by the theory of legitimate peripheral participation [37]. Students who felt less inclined to comment could still participate by reading and voting on other comments. This additional method of participation could explain why students stayed engaged with the digital system for so much longer than observed when writing on the paper notecards.

Students remarked on some negative aspects of the social atmosphere they experienced in the digital system, and there were more off-topic comments in the digital version. We expect this is related to the sense of immediacy the system provides, allowing students to capture their transient thoughts and ideas more easily than when writing feedback on a notecard. When students feel free to respond with what's on their mind, they produce more off-topic comments, but also more comments related to in-the-moment observations such as communication style or points of confusion that might later be resolved.

Providing Organization Support

The digital system offered organization support for students to sort and tag their feedback. While not all students chose to utilize those features, the ones that did found the process valuable to their reflection, which has been shown to be an essential aspect of peer feedback by prior researchers [46].

However, while some teams valued the organization provided by the digital system, the sorting and tagging functionality did not support every team's reflection process. Some teams relied on physically sorting and grouping notecards in a way that could not be captured with the digital system. The digital system had a "Download Responses" button that allowed students to print out their comments, but students did not use this feature.

Handling the Logistics of Peer Feedback

The digital system also provided support for instructors. The system made the logistics of giving and receiving feedback easier, which is a factor teachers value when implementing in-class technology [7]. Teaching assistants did not have to hand out piles of notecards to every chair, team leaders did not have to gather cards from every student, and feedback was never lost in the way a misplaced notecard would be. In larger classes, the paper notecard process would be unmanageable from a logistics standpoint, but the digital system scales.

Effects on Classroom Culture

The instructors commented on the investment students had in their peers' projects. The faculty believed every student was writing notecard feedback, when really only two-thirds of the class participated on paper. Introducing the digital system did not change the number of participants in class.

The digital system did increase the length of time students were engaged in the peer feedback process, but faculty may have been unaware of this difference. When we observed instructors giving verbal feedback before introducing the digital system, they stood at the front of the room facing the presenters, with their backs to the rest of the class. Thus, they could not see when students became disengaged with the feedback process and turned to Facebook or fell asleep.

It is important to note that giving students a specific task to do on their devices kept their attention on course content for a longer period of time than asking them to put their devices away and comment on paper. Students did not report feeling distracted when using the digital system either as a presenter or as a feedback provider. Using a digital system also did not change the quality of the student feedback. The classroom culture around peer feedback did not significantly change because of our two-week system intervention.

Pedagogical Affordances of Digital and Paper

Although students valued the paper feedback more, instructors should choose a peer feedback system that provides the most pedagogically effective approach for a given classroom context. To understand the pedagogical differences between the paper and digital systems, we examined the affordances each provides. Paper provides better tangibility, spatial awareness, and flexibility [26]. The digital system builds on the affordances of social media, including visibility, editability, persistence, and association [62].

In our study, the tangibility of the paper notecards may have made that feedback more salient to students during reflection sessions. Students valued the flexibility to "break the rules" by providing drawings as feedback. Some groups relied on physically organizing the paper notecards, although it created a burden for the instructors and TAs.

The visibility of the digital system influenced students' social behavior, causing some of them to "perform" for the class. Visibility also allowed people to react to each other's comments by voting. The persistence afforded by the digital system meant students could see comments after the fact, promoting a longer length of engagement with the peer feedback process.

Students responded positively to the tangibility of the physical system, even when it created an extra burden for the instructors and TAs. On the other hand, the visibility of the digital system significantly influenced student behavior in both helpful and potentially distracting ways.

Students engaged with the affordances of each system in ways that revealed their needs in the feedback process. Paper sketching indicates that students desire to be expressive (which was further validated by their survey responses) while the use of upvoting suggests students' desire to be social (which was further validated by their "off-topic" comments in the digital condition). Future work can explore how to accomplish both these goals, as well as increase the quality of peer feedback.

Modeling Expert Behavior

In our study, instructors gave more criticism, less praise, and more actionable comments than students. Prior research has noted that students benefit from being more critical of each other when giving feedback [25]. However, listening to an expert give good feedback is not enough for students to improve their feedback skills. Future systems could explore other ways to model expert behavior and encourage relevant, critical, and actionable comments from students. For instance, many recent peer feedback systems explore the use of expert rubrics to scaffold the process [36, 65]. Another idea would be to delegate to a student or a teaching assistant the job of transcribing the instructors' verbal comments, making them visible and available for further discussion in the digital system.

Limitations

The course setup provided constraints for our experiment design. We chose a within-subjects experiment because students could attend both the morning and afternoon sessions of the course. A between-subjects design would have required half the class to use only paper and half to use digital within the same session. This would have made it difficult not only for the instructors to manage, but also for the presenting students to reflect on feedback coming from both notecards and the system. Since both methods (paper and digital) supported anonymity, we were not able to perform within-subjects analysis on an individual basis.

In addition, while we did ask students to rate their feedback as a whole, we did not ask students to individually rate the quality of feedback they received. This information could have been valuable, and we intend to collect student ratings of individual comments in future studies. Future studies should also ask instructors to rate all comments rather than only a subset to better compare to student perspectives.
Our system was deployed for only two weeks, which means that the novelty of the system may have affected our results. Future studies can explore how practices evolve in a digital feedback system over time.

Finally, we chose to compare the digital system to a written feedback process rather than a verbal feedback process because written feedback is given in parallel, asynchronously, and in real time, which is more similar to the digital system than a verbal feedback process. This choice also allowed us to examine specific processes only available in the digital system, such as voting on other students' comments. Future studies could make additional comparative evaluations of in-class peer feedback systems to better understand affordances of different in-class peer feedback approaches. For example, the digital system used in this study or another system could be compared to a verbal feedback process, such as a Q&A period. We also want to compare our in-class peer feedback system to a traditional out-of-class peer feedback system.

SUMMARY

This chapter presented empirical evidence that shows digital methods can be at least as effective as paper at encouraging student interaction during class while offering logistical benefits for instructors. The study found that both methods yielded comments of similar quality and quantity, but the digital approach provided additional ways for students to participate and required less effort from the instructors. While both methods produced similar behaviors, students held inaccurate perceptions about their behavior with each method. This work indicates that students and professors may be overrating the benefits of paper-based feedback and overestimating potential negative effects of technology during class. We also developed design implications for in-class peer feedback systems to enable more expressiveness, to model expert behavior, and to explore different levels of awareness. We also discuss how factors such as classroom culture and student expectations might affect the implementation of an in-class peer feedback system to guide future designers of such systems.

5 STUDY 2 – IMPROVING FEEDBACK QUALITY WITH STUDENT-AUTHORED QUESTIONS

In PeerPresents, feedback receivers write questions that guide feedback providers towards areas that will be most helpful to comment on. Questions are expressed as a set of prompts that help focus feedback on particular issues or concerns, similar to a rubric. For example, if a particular game design assignment asked students to focus on giving players interesting choices, a presenting team might ask peers to comment on how they felt when a particular choice was made in the demo, or what other choices could be added to their current game version.

Rubrics have benefits both for students (e.g. communicating expectations for how their work will be judged) and for feedback providers (e.g. structuring their response). Rubrics can be authored by different stakeholders, such as faculty members, the research team, or students themselves. Faculty- or researcher-authored rubrics may incorporate their *expertise*, while student-authored rubrics require students to *reflect* on what type of feedback they hope to receive. Existing peer feedback systems typically use instructor or researcher authored rubrics rather than student-authored rubrics.

Student-authored guiding questions may help improve peer feedback because they move the activity towards a more effective learning method according to Chi's ICAP model. When students use instructor-authored questions to guide peer feedback, they are participating in a passive learning activity, because students are not manipulating the instructor questions in any way. When students individually write questions to guide peer feedback, they are participating in a constructive learning activity, which ICAP theory suggests would be more effective than passive learning activities. When students collaboratively write questions, they are participating in an interactive learning activity, which ICAP theory suggests would be most effective. This study will investigate the impact of guiding questions when conducted as an active, constructive, or interactive learning activity.

PILOT STUDY

To investigate the possibility of using student-authored guiding questions in peer feedback, I ran a pilot study in the Building Virtual Worlds course where students used PeerPresents for an entire semester. For each of 5 game design projects, student teams would present a demo in class and receive peer feedback. Projects 1, 2, 4, and 5 had an interim presentation one week before the final presentation. Project 3 had only a final presentation. In this pilot study, students were instructed to prepare for the feedback session by discussing the current state of the project with their team, brainstorming 10 questions that would help their peer provide useful feedback, and selecting three of those questions to use in PeerPresents for their presentation.

I collected data on what questions were created in brainstorming, which questions were chosen, log data from PeerPresents, and surveys of student preferences and attitudes towards peer feedback. I also coded the quality of peer feedback as in Study 1.

This pilot found that the quality of student feedback did not improve over time. It also found that feedback receivers felt that most of the feedback they received was unhelpful, even when peers answered the questions the presenting team asked. Students also admitted that they did not enjoy brainstorming questions, and would reuse questions from the first week even if the assignment or stage of the project was very different. Students often wrote general questions that elicited vague, unhelpful feedback.

This pilot study suggests that students do not know how to write appropriate questions to elicit useful feedback from their peers. In this study, students did not receive any training in how to write appropriate questions or in what types of feedback are most useful. This motivates the need to provide students with training on what constitutes high-quality feedback and how to write questions that will elicit high-quality feedback from their peers. In addition, students did not brainstorm collaboratively, which misses an opportunity for interactive learning.
In the next study, I investigated 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?

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. Students used PeerPresents to facilitate feedback exchange [58, 59].

Procedure

Figure 5 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

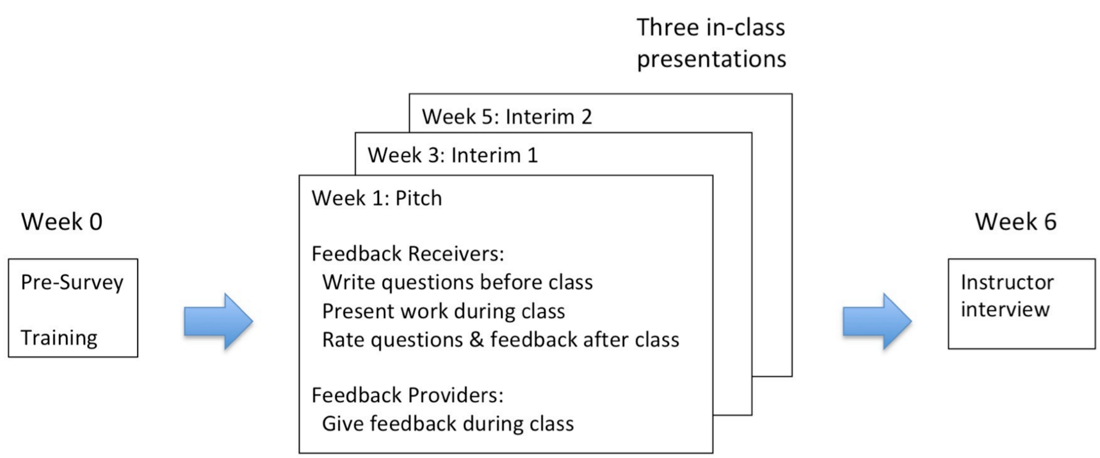


Figure 5: Students completed study tasks before and after each in-class feedback exchange.

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

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

Question Types	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 6: Guiding questions were coded into four types based on what the feedback provider is asked to do.

Analyzing the Type of Guiding Questions

I developed a coding scheme that categorized questions into 4 distinct types. To define these types, a random sample of 50 student-authored guiding questions was categorized using a card-sort method. I then checked the codes against literature that classifies questions for teacher training purposes [6, 23, 47] to see if I 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 6). 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.

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 feed-back providers to change existing game elements; share questions ask feedback providers to provide information or opinions for research and user data. 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. I developed independent coding schemes to address three key qualities of feedback: specificity (the focus of comments), criticality (sentiment), and actionability (grammatical form).

Specificity. I coded specificity of the target and specificity of the insight. I defined the target as the primary game design element mentioned in the comment. I defined the insight as the interpretation or judgment of the target or other extraneous details (see Table 7). 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 this codebook, 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. An insight such as ``looks good'' is general, while ``feels natural playing'' is specific.

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

In the first comment in Table 7, 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

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

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 7, 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. I 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 8 gives examples of feedback with different sentiments.

Grammatical Form. Lastly, I coded feedback comments based on their grammatical form as being either Actionable, Descriptive, or a Question. Two members of the research 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 9 provides examples for each of these three distinct comment forms.

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 9. Analyzing Action-ability: Comments were coded based on their form: actionable, descriptive, or question.

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

Table 10: Students wrote Brainstorm and Critique questions more often than Share and Improve questions.

RESULTS

What types of guiding questions do students write?

As described in the coding scheme, I 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 11).

I 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, I looked at how question type varied at different phases of the project. For values, I looked at student ratings of different types of questions or feedback, since what you value can influence what you write. For process, I 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 6). 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 questions (e.g. "How could we make our educational objectives more specific or clear?") and share questions (e.g. "What is the hardest part of cooking pasta for you?") as more helpful than brainstorm questions (e.g. "What other mini games might we include?") or critique questions (e.g. "How accessible is our current game? Is it too complex, too simple, or just right?"), as shown in Figure 7. 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

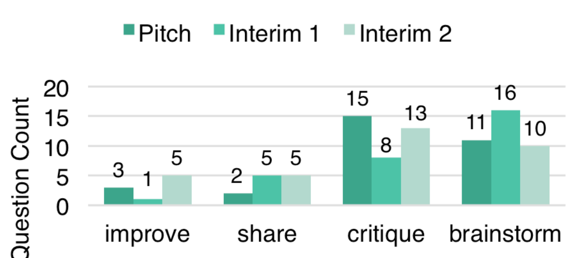


Figure 6: Questions that asked students to critique were the most common during the pitch and interim 2 rounds. Improve questions were the least common.

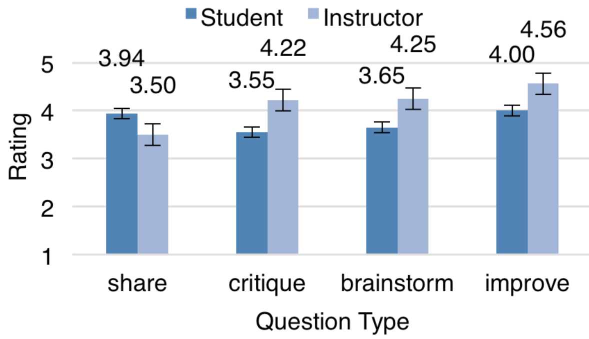


Figure 7: Students rated share and improve questions as more helpful than brainstorm and critique questions. The instructor rated share questions lower than all other question types.

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 value share questions as much as the other types of questions; he rated improve, brainstorm, and critique questions almost equally, as shown in Figure 7.

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 8). 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 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.

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

Table 12: Students and instructors were equally critical when providing feedback.

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

Instructors and students were equally critical.

Students gave a similar amount of praise and criticism compared to the instructor. I 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 11). 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.

I coded feedback into three different types: actionable, descriptive, and question. I 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 12). 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

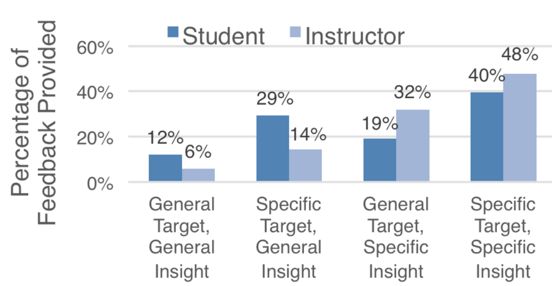


Figure 8: Student feedback focused on specific targets while instructor feedback focused on specific insights.

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 9). 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 9. Examples of improve questions with high response rates were usually phrased as "How can we…". For example, "How can we make this game engaging for longer periods of time than the booth model?" and "How can we better incorporate feedback into our game?" and "How do we make this game less intimidating to people who are unfamiliar with financial products?".

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 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.

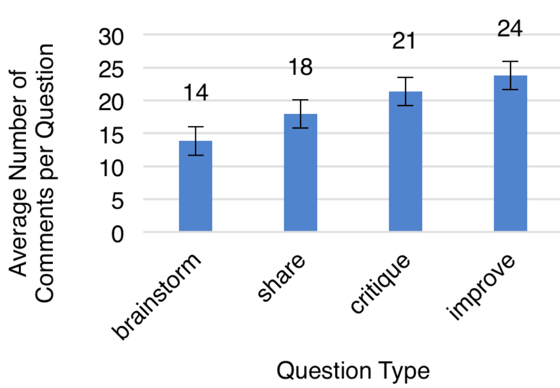


Figure 9: Different types of questions produced different amounts of feedback.

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.

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 [66]. This aligns with ICAP theory, which suggests that the interactive activity that occurred in collaborative teams, but not in individual teams, fostered deeper learning.

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 was 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 [21]. 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.

However, this data does not completely explain why even the students who recognized the value of improve questions still wrote them so infrequently. Perhaps students did not fully identify the difference between improve questions, which ask for a specific suggestion to improve a specific element of the project, and critique questions, which ask for general opinions about either a specific element of the project or the project as a whole. Keeping in mind that students in this study were never told about question types, perhaps future studies that clarify the differences among question types might encourage students to choose improve questions more often.

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.

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.

IMPLICATIONS & 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.

SUMMARY

This chapter 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, I compared a constructive (individual) question-writing process to an interactive (group) question-writing process. This study found that the question writing process influenced the type of questions written and that the questions influenced the quality of feedback received. I 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, student-authored guiding questions may positively impact the peer feedback process and should be explored further.

6 STUDY 3 – SUPPORTING REFLECTION ON AND INTEGRATION OF FEEDBACK WITH A REFLECTION PROTOCOL

This research explores strategies for helping students make better use of their feedback by supporting reflection processes. Reflection can help learners assimilate feedback, and is particularly effective when the learner understands the purpose of the feedback being reflected on.
Using observations of unguided student reflection and existing literature, I developed guidelines for three action-oriented reflection protocols based on the ICAP framework: active, constructive, and interactive. Students in two university game design courses tried all three reflection protocols during the course of a semester. I collected data on student preferences, team dynamics, and quality of peer feedback and student reflection.

DEVELOPING DESIGN GUIDELINES FOR REFLECTION

To develop guidelines for designing an appropriate reflection interface, I first observed student teams in two project-based design courses reflect on their peer feedback on paper. I noticed that student teams did not have a clear strategy for reflecting on the feedback. This lack of strategy often led to teams reading the feedback out loud to each other, then "discussing" the feedback – which was rarely a productive use of time, as the discussion might get off track, a single team member might take over, or many team members might stay silent instead of participating. The discussion rarely identified how the feedback could be used to improve their project. For the teams who did have a strategy, students would often simply count the number of times comments were repeated. The quality of team discussion in these cases was not any better than teams without a strategy.

There are two major problems with the reflection I observed. One problem is the lack of interaction during reflection. Currently, student reflection would be categorized as an active learning activity in Chi's ICAP framework, because students are taking the information they receive from the feedback and manipulating it to count how often each idea is expressed. However, team reflection has the opportunity to be a constructive or even interactive learning activity, which theory suggests would increase learning gains. For example, in the SWoRD system created by Cho & Schunn [13], feedback receivers reflect on feedback by providing a 1 to 5 star rating for each comment they receive. This feedback-focused reflection could be categorized as a constructive learning activity, because students must generate a rating for each comment. This metacognitive process helps deepen their learning.

Second, students are not able to use feedback to iterate on their projects. Prior work on feedback suggests that iteration-focused reflection for individual designers, where designers consider feedback in the context of what next steps they will take in iterating their design, is more effective than unstructured reflection [64]. If students reflect on peer feedback by collaboratively generating a list of action items for their project, their reflection activity would be considered interactive learning and could help them use feedback to iterate their work more effectively.

Using both the ICAP theory and the prior work on iteration-focused reflection, I designed three reflection protocols. The first protocol is the Active learning protocol – students read the feedback and discuss without guidance, just as we observed students already doing. The other two protocols have students generating a list of next steps for the project (iteration-focused); one protocol has students generating lists individually (Constructive learning) while the other has students generating lists collaboratively (Interactive learning).

Method

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, I 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 what research says are effective characteristics for feedback: specific, critical, and actionable. 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 courses, students participated in the study after each interim presentation of their work in class; we did not conduct the study when students presented a final version of their work. Figure 10 and Figure 11 below show how the study fit into the overall course. In the graduate course, each project lasted two weeks, so student 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.

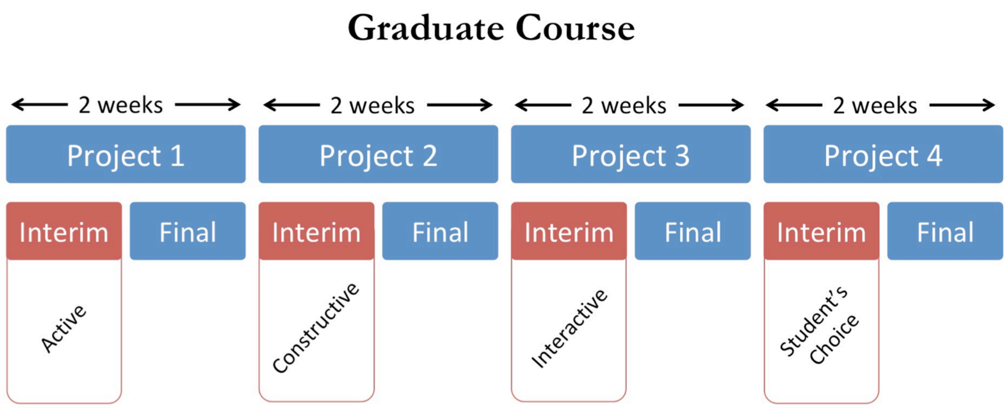


Figure 10: The graduate course had four projects, each lasting two weeks. Students participated in the study every other week, for four total interventions.

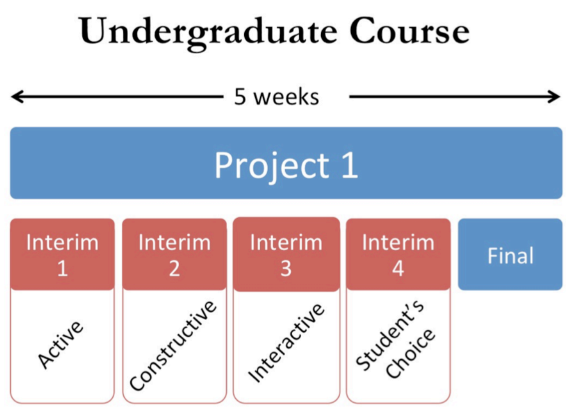


Figure 11: The undergraduate course had one project that lasted five weeks including the final. Students participated in the study every week, for four total interventions.

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. After class, teams would meet with a researcher at their chosen time slot. During this meeting, the researcher would obtain consent, guide the team through a reflection protocol, and give them a check-in survey.

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 of the in-class feedback exchange. Individuals who participated were compensated with a $5 Amazon gift card for each meeting they attended.

Reflection Protocols

I designed three reflection protocols for students to follow based on a literature review of iteration-focused reflective feedback practices (see Table 13). At the start of all three conditions, students have 10 minutes to read the feedback their team has received. As they read each comment, students rate how helpful that comment was on a scale of 1 to 5. The activities for the remaining 15 minutes varied by condition.

Active	Constructive	Interactive
• 10 minutes – silently read & rate feedback	• 10 minutes – silently read & rate feedback	• 10 minutes – silently read & rate feedback
• 15 minutes – team discussion	• 5 minutes – individually write next steps	• 10 minutes – team discussion
	• 10 minutes – team discussion	• 5 minutes – collaboratively write next steps

Table 13: Students were guided through one of three reflection protocols during each reflection meeting. All three protocols took 25 minutes, beginning with a reading period and including at least 10 minutes for team discussion.

In the Active condition, after students finish 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 think are the most important actions to take on this project. After students write their list, they have 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 at the end to collaboratively write a list of next steps that the entire team agreed on. 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 system design of an effective reflection interface.

Check-In Survey

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. The 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.

Measures and Analyses

Quality of Peer Feedback

The quality of peer feedback was measured in two ways.

1. **How often do students provide specific, actionable, or critical feedback?** As in prior studies (see Chapter 5), researchers coded for three qualities of successful peer feedback: specific, actionable, and critical. These codes are listed in Table 14 below, and described in detail in Chapter 5.

2. **How do students perceive the peers who provide feedback?** I also analyzed data from the check-in survey from Likert-scale questions asking students their opinions about the peer feedback process and the peer feedback provider.

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

Table 14: Researchers coded for three qualities of successful feedback: specific, actionable, and critical.

Quality of Student Reflection

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

1. **How do students rate the feedback they received?** Students rated the feedback they received on a scale of 1 to 5. I analyzed average student ratings and if ratings changed over time.

2. **Do student ratings align with expert codes?** To answer this question, I compared the researcher codes for the three indicators of feedback quality (see Table 14 above) 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 rating based on the codes to compare to student ratings. I 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). I 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. However, because I didn't want to penalize cards that contained "fluff" statements that might have a very low score, I removed all comments that had a zero or one score before taking the average for the notecard. This meant statements like "Good job!" or "Your music was awesome!" would not arbitrarily bring down the overall notecard score.

3. **Do student ratings align with other team members?** I measured team disagreement by calculating the standard deviation among team member's ratings of individual notecards. I

report the average standard deviation across all teams for each interim presentation, and analyze if ratings change over time.

Student Preferences

I measured student preferences in two ways: students' choice for which protocol to follow and students' expression of preference.

1. **Which protocol do students choose?** I counted how many teams choose each reflection protocol during the final "choose your own adventure" week.

2. **What do students like and dislike about each reflection protocol?** I categorized how students responded to survey questions about their likes and dislikes for each reflection protocol using a grounded theory approach to identify common themes.

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

Quality of Peer Feedback

Researchers coded feedback for specific, actionable, and critical as indicators of feedback quality.

Students usually provide specific feedback, but are much less actionable or critical.
Students were almost always specific when giving feedback. They referred to a specific target in 83% of their feedback, and provided a specific insight in 81% of their feedback. Students gave actionable feedback in 33% of comments and critical feedback in 38% of comments.

Students felt peers were qualified to give feedback, but perhaps lazy.
We also surveyed students after each reflection meeting about their perceptions of feedback quality, asking questions like ("The feedback I received provides helpful suggestions for improving my work" and "The feedback I received is fair" & "is relevant"). Student responses to these questions were positive and did not change over time.

Student ratings of the feedback provider were consistent across conditions, although one student commented that "I believe my peers are qualified and capable, but on some occasions the feedback cards do not feel like an active attempt to give good feedback, but rather the[m] going through the motions to meet the requirements." We should have asked about this sentiment more directly.

Quality of Student Reflection

Students consistently gave below-average ratings to the feedback they received.
Graduate student ratings of feedback quality averaged 2.7 on a scale of 1 to 5. Student ratings of feedback did not change significantly over time (see Table 15).

Student Ratings of Feedback Received	
Project 1	2.82
Project 2	2.76
Project 3	2.70
Project 4	2.63

Table 15: Student ratings of feedback they received averaged 2.7 out of 5.

Student ratings did not align with the amount of specific, critical, or actionable feedback.
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 ($R^2 = .017$), but a graphical representation of the data reveals a potential trend (Figure 12). For example, if a comment was specific in insight, the insight might be helpful or unhelpful, as evidenced by the variety of ratings for comments that are 100% specific. But, if the comment was less specific, it was less likely to get a high rating from students, as evidenced by the lack of data points in the upper left quadrant of the figure.

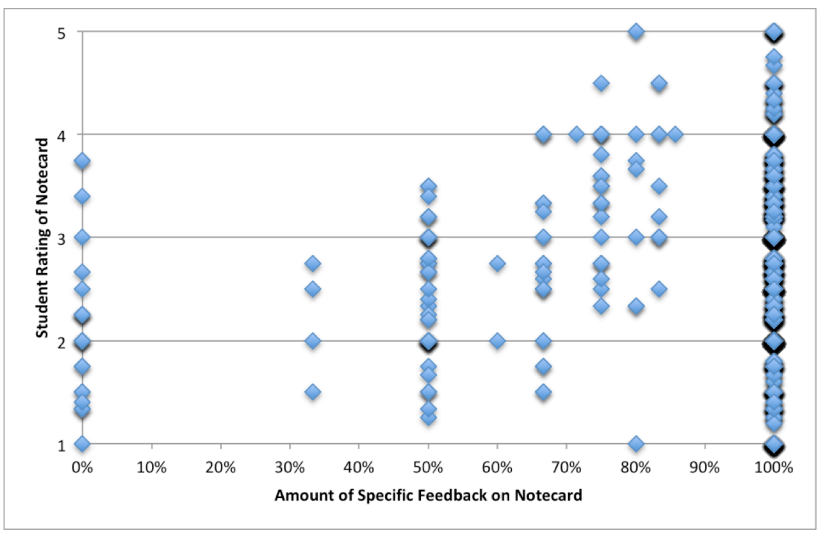


Figure 12: Specific comments could be rated high or low, but less specific comments were less likely to get a high rating from students.

Students gave higher ratings to longer comments.
As shown in Figure 13, 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"

and "Very thorough, complete experience. During the early narration, maybe talk a bit about the sun being mean or whatever" were both highly rated comments with less than 150 characters.

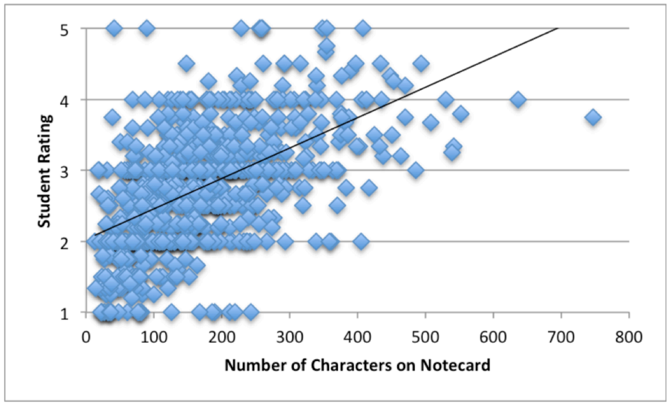


Figure 13: Student ratings aligned with comment length.

Team disagreement when rating feedback decreased over time.
Team disagreement when rating feedback decreased slightly over time (see Table 16). In the undergraduate course, students were working with the same team throughout the semester. However, in the graduate course, students were working with a completely new team for each project. Thus, the increase in team alignment is not due to getting more comfortable with a consistent group of students.

	Standard Deviation of Student Ratings
Project 1	1.12
Project 2	1.08
Project 3	1.05
Project 4	0.98

Table 16: Student disagreement when rating feedback decreased over time.

Student Preferences

We analyzed student responses to the check-in survey to identify the four most common areas where students expressed preferences. We also identified student goals for the reflection meeting, and analyzed how many teams chose each reflection protocol during the "choose your own adventure" week.

Students preferred the Interactive reflection protocol.
In the "choose your own adventure" week, most teams (11 out of 14) preferred to follow the interactive protocol. Only three teams choose 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 would typically do in a reflection meeting without researcher intervention.

The reflection protocol influenced student goals.
When commenting on their likes and dislikes, students also expressed their goals for the reflection process and how the activity did or did not help them achieve these goals.

- Sensemaking & Reflection
- Iteration
- Prioritization
- Team building
- Efficiency

The frequency of some of these goals changed depending on which conditions students were in (see Figure 14). As students progressed from active to constructive to interactive, they focused less on just reading and understanding the feedback (sensemaking and reflection) 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.

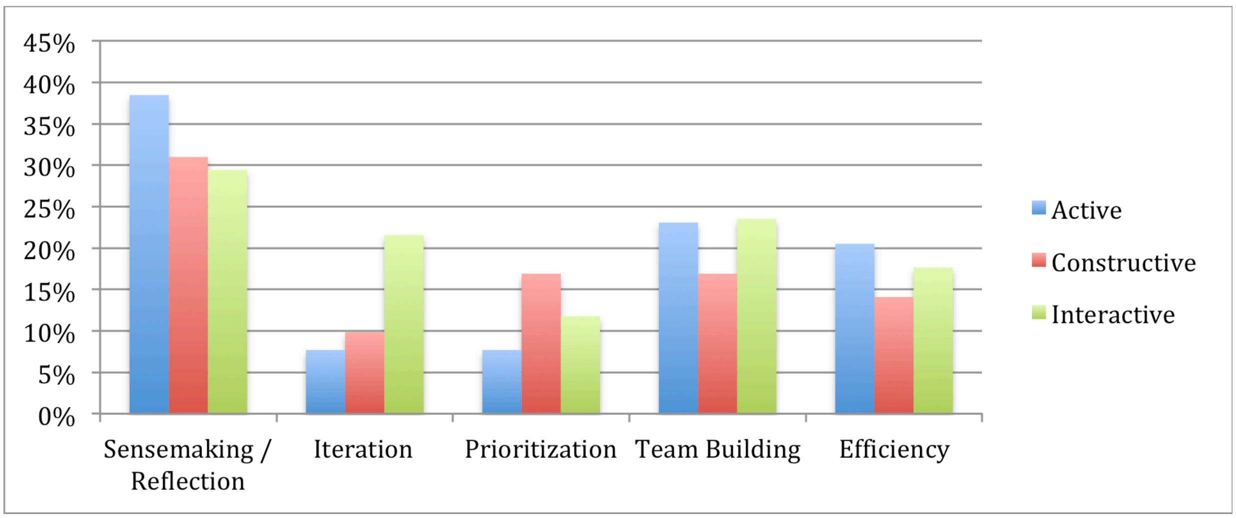


Figure 14: The goals students expressed for the reflection activity changed based on the reflection protocol. Students focused less on sensemaking and more on iteration as the reflection became more interactive. Students also highlighted a tradeoff between prioritization and team building in the constructive condition.

Sensemaking: 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", "[gave] individuals more time to think before discussing", and provided time to "come to my own conclusions before sharing". Students felt this made discussion more efficient because they didn't have to "waste time thinking of what to say next".

Iteration: Students had concerns about feedback quality and quantity.
Students described feedback quality in terms of how they could use the feedback to iterate or improve their project. Students often commented that the quality of peer feedback was low, describing feedback as "not helpful", "not useful", "unclear", and "shallow", although the

percentage of students who voiced this complaint decreased as the semester continued (23% in the Active week, 16% in the Constructive week, and 8% in the Interactive week). A smaller contingent of students said the feedback was helpful, most often mentioning the benefit of getting diverse perspectives on their work when iterating. Some students felt they were getting too much feedback, and suffered from "information overload" since "dealing with all of [the feedback] could be tiring".

Prioritization: Students felt ratings helped them 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". 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". Students also expressed a feeling of information overload after too many ratings.

Prioritization: Students felt next steps helped them prioritize.
When written individually, students felt the list of next steps helped them prioritize their own thoughts. They also used the lists to compare priorities across team members and to identify areas of agreement. They felt this activity made the discussion more focused and productive. When written collaboratively, students felt they were able to identify their priorities as a group. However, one student mentioned the list-making process was taken over by a single team member, and two others expressed a direct preference for individual list-making.

Team Building: Students remarked on the open nature of the discussion.
Many students valued the open and unrestricted nature of the discussion. They appreciated being able to talk about whatever they wanted, and they described the discussion as "open", "honest", and "clear". However, other students wanted the researchers to provide more direction to prevent the discussion from getting off track, saying "I thought we would have help from [the researchers]" and "[I disliked that] we still had to self-determine the direction of the discussion". A few students mentioned a design to keep better records of their team conversation.

Team Building: Students wanted reflection to be a team building activity.
A common theme was the desire to include the whole team in feedback discussion. Students commented negatively about the activity both when "not all of the team was here" for the meeting and when individual team members were either "monopolizing" the discussion or if a "quieter member of the group" was saying nothing at all.

Efficiency: Students noticed the time constraints.
Students had mixed opinions about the time limit researchers placed on the activities. Some students 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" and "[the] time limit helps [us] summarize feedback quickly". 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" 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". 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".

Team Dynamics

The videos of team meetings revealed multiple strategies teams use when reflecting on feedback and dealing with or avoiding disagreement on the team.

Teams focused on suggestions.
Teams focused on suggestions they found in the feedback. For example, Team A started their feedback discussion by having each team member read out loud their favorite notecards that contained suggestions for solving the problems in their current game iteration. Students also judged feedback quality based on the number of suggestions. Team D mentioned they wanted more suggestions from the feedback, and Team G decided their feedback was not very useful because it didn't have a lot of suggestions.

Instructor feedback trumps peer feedback.
Teams sometimes disagreed about whether to follow peer feedback or instructor feedback. For example, in Team A's meeting, one team member argues for keeping a section of dialogue because the peers really liked it, but another team member argues for removing the dialogue because the instructor didn't like it. The students did not offer any additional reasons for their side of the argument. Team B had a similar argument between peer and instructor feedback (described below as a Compromise). The instructor feedback often became the focus of discussion, as the instructor's opinion was considered to be more important than peers'.

The most popular 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.

Teams vote on what feedback to discuss and implement.
When teams were in the Constructive condition, they would often use their individual lists as a way to "vote" on what to discuss next. For example, Team F and Team A started their discussion with a next step that every team member had on their list, then moved on to discuss less popular next step items. Other teams would "vote" on what feedback to implement, even if the team opinion disagreed with the feedback. For example, in Team B's meeting, one team member pointed out a suggestion from the feedback, but because the other two team members in the meeting felt it was ok "just to ignore it", the team did not implement that comment.

Teams seek compromise.
Teams would often seek a compromise, either between two pieces of conflicting feedback or between two team members' ideas. For example, when Team B discussed a transition point in their game, one person noted the negative feedback from peers while another person noted the positive feedback from the instructor. A third team member suggested that perhaps the transition was conceptually ok, but not visually ok; the team was able to compromise by adding a tunnel visual to the existing transition.

Strong team leaders can hinder team communication.
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." Other students remarked that one team member might take over a stage of the reflection meeting rather than soliciting everyone's opinions.

Discussion

Students analyze feedback quality using different metrics than experts.

Experts refer to specific, actionable, and critical as markers of high quality feedback. 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 to others. While students did give specific feedback often, 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. 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 influence student goals and team success.

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. To address this need, I designed a peer feedback skills inventory base on the Team Dynamics data to identify and explore strategies students use when reflecting on peer feedback. The design and pilot of this inventory is outlined in Chapter 7.

Reflecting on feedback can be a team alignment activity

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 support this by not only framing reflection as a team activity for students, but also by providing 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. First, 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. Second, this study revealed both positive and negative implications of enforcing time limits on discussion. 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. Third, 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 [40]. 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. Finally, 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. 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.

Limitations

One major limitation of this study design 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 change.

7 PEER FEEDBACK SKILLS INVENTORY

Skills inventories are used to help identify which skills in a particular domain an individual already possesses and track change in skill development over time. For example, the Force Concept Inventory [28] asks students a series of questions about Newtonian concepts. One answer choice for each question represents a "correct" Newtonian concept, while the other answer options align with "commonsense alternatives" to Newtonian physics. When students answer incorrectly, they reveal which misconception they are operating under. This allows educators to not only diagnose student misconceptions but also alter their instruction to address those misconceptions in the future.

Instructors are often familiar with the common misconceptions students experience in their own domains. However, effective peer feedback exchange requires both domain skills and peer feedback skills. Most domain instructors are likely unfamiliar with the specific misconceptions or common pitfalls student experience when trying to improve their peer feedback skills. Thus they are at a disadvantage when trying to instruct students on best peer feedback practices.

Adding even more complexity is a key difference between concrete skills like solving physics problems, where there is a consistent right answer, and dealing with interpersonal issues and team dynamics during feedback exchange, where there are often situations in which there is no right answer. However, there might be consistently helpful strategies or consistently unhelpful strategies in dealing with peer feedback exchange. Before strategies can be evaluated, they must first be identified.

This chapter describes my work to identify common strategies and misconceptions students experience when using peer feedback skills and design a set of questions that could be used as a skills inventory for peer feedback. While there are many phases of the peer feedback process, I focused on the reflection phase for this skills inventory, because students often have the most trouble and the least guidance during the reflection phase. This work seeks only to describe student strategies and perceptions; we do not attempt to evaluate the strategies that are identified.

Designing the Peer Feedback Skills Inventory

In Chapters 4 and 5, I observed student behavior in unstructured team reflection meetings. In Chapter 6, students completed reflection tasks during structured team meetings. From both structured and unstructured meetings, observational data revealed a set of student goals for dealing with feedback. Students valued efficiency – they wanted the meeting to run smoothly and the discussion to stay on track. Students sought sensemaking and reflection – they wanted to read and understand the feedback their team had received. Students valued prioritization – because they were receiving a large amount of feedback and had a limited amount of time left on their project, they wanted to identify which feedback was most important and what the team felt was the highest priority. Students focused on iteration – the feedback was meant to help them improve their project. Students sought team building – the reflection activity was done as a team, and the students wanted to feel their team was on the same page after reflection.

When students followed structured reflection protocols, the reflection strategies students used influenced how students talked about their goals for reflection. For example, when students followed the Active protocol in Chapter 6, they talked about sensemaking, or understanding the feedback, as a primary goal. But when students followed the Constructive and the Interactive protocols, they focused more on iteration, or using the feedback effectively, rather than merely understanding the feedback. Students also expressed a tradeoff between prioritization and team building, depending on whether the reflection meeting had an individual list-making activity. The activities students completed during reflection influenced student mindset during reflection.

This work explores if and how other strategies influence student mindset during reflection. First, I identified distinct steps in the reflection process to serve as a framework for the feedback strategies. Second, I analyzed the data from team reflection meetings in Chapter 6 to identify strategies students used during each step of the reflection process. Then I created a set of questions for the skills inventory designed around a scenario that asks students to identify which strategies they prefer and explain why. We believe this is a first step towards a peer feedback skills inventory that can identify how and why students are reflecting on feedback and if their strategies are effective.

Defining the Reflection Process

I broke down the peer feedback reflection process into distinct steps. When students are given a collection of feedback comments, they first need to summarize the feedback they have received. Next, students have to deal with conflicting or contradictory feedback. Then students need to identify appropriate next steps for their project based on the feedback.

Identifying Strategies

I identified specific strategies that I have observed students using to complete each of the reflection tasks. These strategies are listed in Table 17 and described in detail in the following paragraphs.

Summarization Strategies	Conflicting Feedback Strategies	Next Step Options
• Count each idea	• Compromise	• New features
• Suggestions only	• Choose most popular	• New look & feel
• One theme only	• Team vote	• New idea
• "Everything is fine"	• Ignore both	

Table 17. We identified four strategies for summarizing feedback, four strategies for dealing with conflicting feedback, and three areas of focus for the next step.

Strategies for Summarization

When students summarize their feedback, I have observed four common strategies. The first strategy is to count how many times each comment is repeated. Students might rewrite paraphrased comments and keep a tally of how often each one appears, then focus on the most popular comments. However, students sometimes heavily paraphrase comments to combine ideas in ways that manipulate the final tally. A second strategy is to focus on the concrete suggestions given by feedback providers, and ignore feedback that is not a suggestion. This allows students to focus on clearly outlined action items, but helpful feedback that is not phrased as a suggestion might get lost. A third strategy is to pick out one common thread or theme from the feedback and ignore the rest of the comments. This could be useful if that one topic must be addressed before the project can proceed, but it also might over-simplify the breadth of the feedback. A final strategy is when students ignore or over-summarize feedback, saying "the feedback is positive; our project is fine – we can just keep going as planned".

Strategies for Conflicting Feedback

Students almost always receive contradictory feedback when a large volume of feedback is collected. I identified four strategies students often use. First, students try to find a compromise and incorporate parts of each side of the conflict. Second, students count how many feedback providers mentioned each side and go with the most popular. Third, the team could take a vote and go with whichever side they prefer. Fourth, the team could decide to ignore both pieces of feedback.

Strategies for Defining Next Steps

In game design projects, there are typically three areas in which next steps could be completed: adding new features to the existing project, changing the look and feel, or refining the idea. It's less important to identify which of these three areas students choose and more important to identify *why* that area was chosen. Particularly, did the decision come from something in the feedback, or did it come from the student's personal opinion, or from some other factor?

Creating a Scenario

Finally, I created a scenario as a framework for asking questions about student strategies in a concrete example context. The scenario text is:

> "Imagine you are in a UX course working on teams. Your fictional client is the committee organizing an upcoming conference, and they have asked you to help improve their conference website to make it easier for participants arriving at the conference. Your team wants to design a website to help student attendees find a roommate before their arrival so they can share housing costs. Imagine you and your team presented a click-through demo of the site in class. After this presentation, you have three weeks to implement your design."

Students were also given a list of 9 feedback comments (see Table 18). The comment list included positive and negative opinions about the idea and the look and feel, as well as different ideas for new features. This diversity ensured there were contradictory comments for whichever area students decided was most important to pursue.

Sample Feedback	Area	Sentiment
C1: I think the client is less concerned with helping students find housing; they want to help faculty attendees find the right room for their talks.	Idea	Negative
C4: Why not just use Facebook for this instead?	Idea	Negative
C5: Helping more students attend a conference is really important! Conference committees need help with this. Cool idea.	Idea	Positive
C6: The concept works but your interface design looks outdated.	Look & Feel	Negative
C7: You need to make the layout more mobile-friendly.	Look & Feel	Negative
C3: Great work! Your website is beautiful	Look & Feel	Positive
C2: I really like that you're helping students find a place to stay. Maybe include an option for them to find a roommate after they already have housing as well as finding a roommate first	Features	
C8: Have you thought about including more search filters beyond just cost and dates?	Features	
C9: Awesome!		Positive

Table 18: A list and classification of the feedback provided for the scenario.

When creating the scenario questions, each strategy had a multiple-choice question where each answer choice aligned with one of the identified strategies. Then the student was asked why they picked that strategy as a free response question. The skills inventory did not assume "correct" answers for the strategy questions; it is more concerned with probing students to recognize that they are using a strategy, explain why that strategy was selected, and potentially reveal their awareness of pros and cons of that strategy choice. Table 19 outlines the alignment between answer choices and feedback strategies for our scenario.

Summarization Strategies	Which summary do you feel is the most helpful?
Count each idea	Four people thought it was great. One person thought the idea wouldn't appeal to our client. One person thought Facebook could do this instead. Two people thought we should change the interface design. One person thought we should include more search filters.
Suggestions only	The feedback suggested that we include an option for them to find a roommate after they have housing, update the interface, make it mobile friendly, and include more search filters.
One theme only	The feedback suggested that we reconsider our idea - the client probably cares more about faculty than students, and students could just use Facebook for this instead.
"Everything is fine"	The feedback gave us a few suggestions, but most peers really liked our idea. We're on the right track.
Contradictory Feedback Strategies	**Pick one of the following strategies you could use to address this contradiction with your team:**
Compromise	Find a compromise between the two - try to incorporate both perspectives.
Choose most popular	Count how many peers agreed with each comment - go with the most popular.
Team vote	Let the team vote on which one to do - we have the best perspective on this project.
Ignore both	Ignore both pieces of contradictory feedback - they cancel each other out.
Next Step Options	**Based on the feedback your team received, what do you think is the best next step for your team to pursue on this project?**
New features	Add additional features to our existing idea (additional search filters, etc.)
New look & feel	Change the look and feel of the interface design
New idea	Refine our idea - a roommate finder isn't going to appeal to the client

Table 19: Each answer option in the scenario aligns with one of the identified strategies for feedback.

Piloting the Peer Feedback Skills Inventory

Method

Students in Study 3 also took a pilot version of the peer feedback skills inventory. The survey was voluntary, taken online through a Google Form outside of class time. The survey took less than 15 minutes to complete. Students who completed the survey were compensated with a $5 Amazon gift card. Students were offered the survey midway through the semester, so they had multiple opportunities to receive and reflect on peer feedback before taking the survey.

Results

Which strategy would you use to summarize feedback, and why?

Of the four summarization strategies, 58% of students chose to list the suggestions, and 40% of students chose to count how often each idea is expressed. Only one student chose the "everything is fine" summary, and no students chose to prioritize one theme only. Students who chose to list suggestions expressed a desire to focus on "actionable changes" and the "specific things that can be fixed". Students who chose to count each idea felt this summary was "all encompassing", "unbiased", and the "most objective" because it mentioned everything that was said.

Which strategy would you use to resolve conflicting feedback, and why?

Of the four conflicting feedback strategies, 61% of students chose to find a compromise, 23% chose to count to identify popular comments, 14% chose a team vote, and only 1 person felt it was ok to ignore both if they were minor contradictions. Students who preferred compromise felt it was "important to consider every opinion" that the feedback providers expressed, and students who preferred popularity felt that focusing on "pleasing the most" users would allow them to "target a larger audience". In contrast, students who preferred a team vote felt strongly that "there's no right answer" and "we [the design team] know best" because "it's our project".

Which next step would you pick, and why?

Of the three next step options, 54% of students preferred to add new features, 30% wanted to change the look and feel, while only 16% wanted to redesign the idea.

Those who preferred new features felt that the feedback providers gave them new feature ideas, so those ideas should be implemented. They also expressed that functionality comes before aesthetics. One student felt that in order to iterate and get more feedback, they <u>had</u> to "add more" to their project. Another student expressed how disappointing it is to redo work as an argument against refining the idea.

Those who preferred changing the look and feel also felt that the feedback providers had suggested these improvements. They also expressed that aesthetics make it more functional, that there's "no need for more features if they aren't accessible", and that "I'm a UX designer and this is what UX designers do". One student expressed that using aesthetics to distinguish the project from Facebook might alleviate some of the client's concerns with their idea.

Those who advocated for changing the idea had one unanimous reason: the client is the most important perspective. If the client's goals are not being met, then we must change our design to meet those goals.

Do strategies influence each other throughout the process?

Summarization strategy did not impact strategy choice for resolving contradictory feedback. We found that how you summarize does not influence how you resolve contradictions (see Figure 15). Students who summarized by counting each idea were just as likely to resolve contradictions in the same ways as students who summarized by focusing on suggestions.

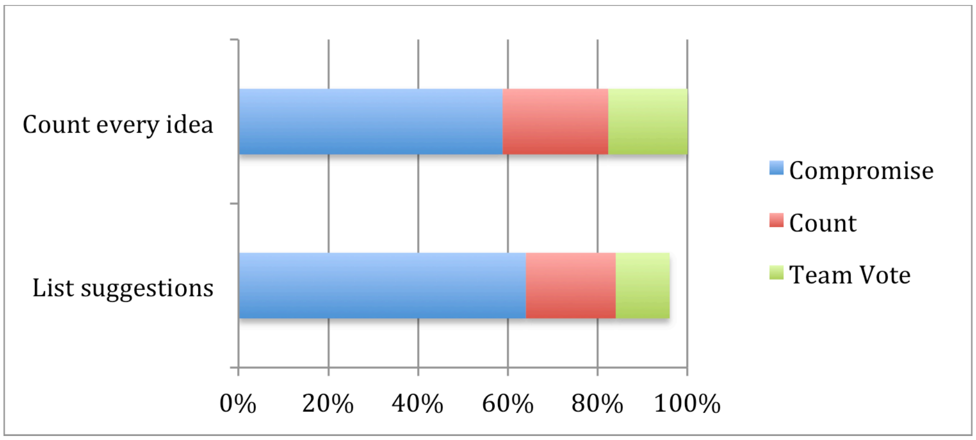


Figure 15: Summarization strategy did not impact strategy choice for resolving contradictory feedback.

Students who focused on suggestions were less likely to consider changing their idea. The next step you pick might be influenced by how you summarize (Figure 16). Students who focused on suggestions would pick either new features or look and feel changes, but would not change their idea. This could be because the suggestion comments listed actionable features or actionable look and feel changes, but not the more abstract feedback that indicated the idea might need refining.

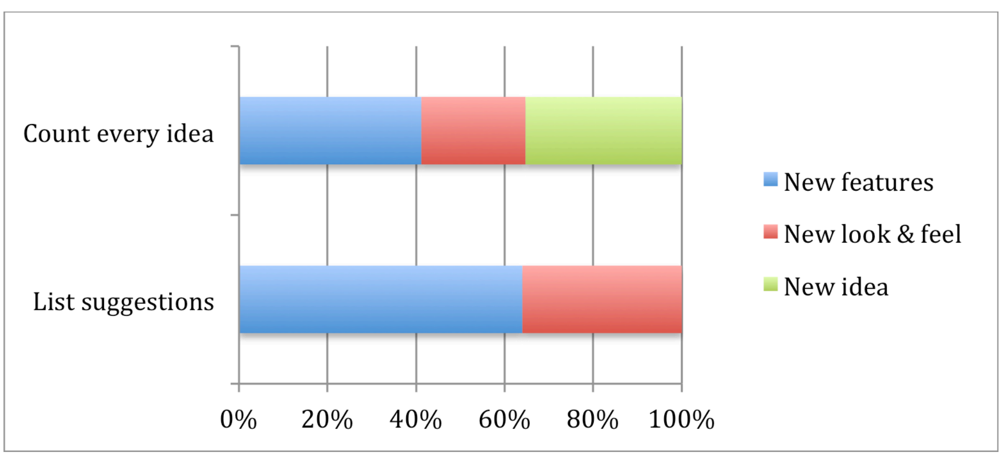


Figure 16: Students who focused on suggestions were less likely to consider changing their idea.

Students who valued compromise were less likely to consider changing their idea. The next step you pick might also be influenced by how you resolve contradictions (Figure 17). Students who preferred to compromise were much less likely to want to change the idea than students who resolved contradictions another way. This might be because it is difficult to compromise on the argument of "should we change the idea or not" – there's no going halfway. But finding a compromise between two new feature choices or two ideas for changing the look and feel is much more possible.

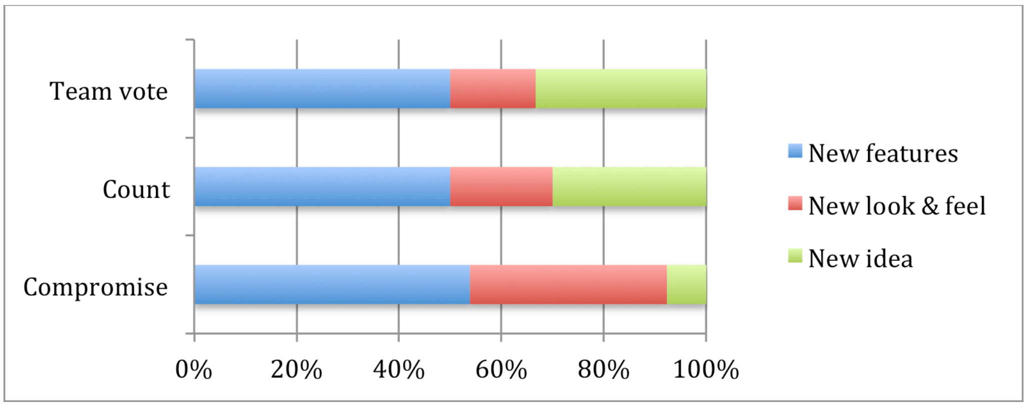


Figure 17: Students who valued compromise were less likely to consider changing their idea.

Discussion

Interventions are most effective at the moment when students interact with the feedback.

The data reveals that the strategies students use when they summarize feedback and deal with conflicting feedback impact their choice of next step for their project. Students who focused on suggestions or valued compromise were much less likely to consider changing their project idea.

This data implies that the moment when students interact with the feedback – when they summarize and resolve contradictions – is the most important moment to intervene. It is at this moment when students are making decisions about what to do next on their project. Waiting to intervene until after students have digested the feedback is likely too late to impact their mindset and goals.

Feedback strategies impact team dynamics and project iteration.

Student goals for the project were impacted by their chosen strategies for reflection. This echoes the findings from Chapter 6, which found that a change in the reflection protocol caused a change in students' self-report of their goals for the reflection meeting. This suggests that how a student reflects is significant, and that a reflection method has the potential to impact not only team dynamics but also the team's project iteration – both of which are factors that impact student performance and learning in game design courses.

This finding also suggests reflection strategies as a framework for future interventions. Changing how students reflect, both at the individual and the team level, could impact student performance. For example, it could be that if students were encouraged to use the same strategy when summarizing or when dealing with conflicting feedback, that they might have lower rates of team disagreement or higher quality team discussion. Future investigations could also explore measuring the success of each reflection strategy to come up with recommendations. Even if there isn't a "one size fits all" strategy for reflection, future work could analyze how factors such as the phase of the design process, team size, team leadership, and team experience impact which strategy might be most effective for students.

Design implications for future systems

Future systems could use these new findings to improve students' learning. Systems could deploy interventions during the feedback interaction phase of the reflection process, rather than after the student has digested their feedback. Systems could also design interventions around reflection strategies, and highlight which strategies students are using. For example, the PeerPresents reflection interface could encourage students to write a summary of the feedback within the system, then use natural language processing to detect which summarization strategy each team member is using. The system could even suggest and explain alternative strategies. Future systems could also help collect the data needed to evaluate which strategies are most effective. Following these implications in future system design could help change the mindset of students in ways that might align student goals with the team, and align team goals with the goals of the discipline.

Future Work

Validating the peer feedback skills inventory was out of scope for this dissertation. However, there are several ways the validity of the skills inventory could be evaluated. Future work could analyze how strategies affect a student's ability to evaluate feedback. In Chapter 6, students rated the feedback they received. The accuracy of these ratings, as a measure of student ability to recognize feedback quality, could be compared to strategy choices. For example, maybe students who use the "count every idea" summarization strategy are less accurate raters of feedback quality than students who focus on suggestions. Future work could also analyze if students who prefer a certain strategy also value one particular feedback quality over another. Maybe students who seek compromise value neutral feedback more than criticism, for example. Future work could also look at whether and how feedback skills change over time. We would expect that students would get better at feedback skills with practice, and that teams would align towards common goals over time, so future analyses could investigate whether this expectation is supported by the skills inventory.

This skills inventory focuses on feedback skills during the reflection phase. However, there are other skills that students use in the preparation and exchange phases that are not reflected in this inventory. For example, future work could identify what skills students need to give high quality feedback and design a skills inventory to evaluate what knowledge or strategies students need to improve during the "feedback provision" phase.

8 Conclusion

Summary

In this dissertation, I used interactive learning activities to address open challenges of peer feedback systems. I first conducted a literature review to identify three major open challenges with existing peer feedback systems. Existing systems struggle to help student **engage** in the feedback process, **improve** the quality of feedback they give, and **reflect** on the feedback they receive. Those three open challenges serve as the organizational framework for this dissertation. The literature review also identified interactive learning activities, where students collaborate with others to produce new knowledge, as a potential solution for the open challenges this work addresses.

After identifying open challenges, I developed a framework for thinking about peer feedback as an end-to-end process, starting with how students prepare for feedback exchange, going through the feedback exchange, and ending after students have reflected on feedback. This framework suggested appropriate opportunities for when to address each of the open challenges (see Table 1). Students can participate in activities to **improve** their feedback quality as they prepare for feedback exchange. During feedback exchange, students participate in activities that help them fully **engage** in the process. Afterwards, students participate in activities that help teams **reflect** on feedback they receive.

After developing a framework for interventions, I then developed a novel peer feedback system to serve as a platform for those interventions, as described in Chapter 3. PeerPresents is designed for in-class peer feedback on unfinished projects. It supports student activities before, during, and after feedback exchange that address the open challenges of peer feedback systems.

After developing a platform for interventions that aligns with the framework, I finally designed the learning activities that would help solve each of the open challenges for digital peer feedback exchange.

To address the challenge of engagement, I moved feedback exchange into the classroom and designed an activity for feedback providers to view and vote on the comments others had written. As described in Chapter 4, when we compared digital to paper-based feedback exchange, students using the digital system engaged in feedback exchange longer than students who exchanged feedback on paper.

To address the challenge of improvement, I designed two "guiding questions" activities for feedback receivers. Receivers write questions before feedback exchange either individually or collaboratively with their team. We found there were four types of guiding questions written by students: Improve, Critique, Brainstorm, and Share. As described in Chapter 5, the type of guiding question impacted feedback quantity and quality, and students who wrote questions collaboratively were more likely to choose more effective questions.

To address the challenge of reflection, I designed three "reflection protocol" activities for feedback receivers (see Table 13). When teams meet to discuss their feedback, they first spend time reading and rating the individual comments. Then, they have time for team discussion and a chance to write next steps for their project either individually or as a team. As described in Chapter 6, the reflection protocol influenced student goals for the team meeting and team dynamics. In Chapter 7, I further explored reflection strategies with a peer feedback skills inventory. The inventory revealed that strategies students use when summarizing and resolving contradictions during feedback reflection impact their team dynamics and the next steps they choose to take on their project.

Key Takeaways

Interactive learning activities are an effective way to address challenges of digital peer feedback exchange.

This work shows that harnessing the power of interactive learning activities can be an effective way to address issues like student engagement in feedback exchange, quality of peer feedback, and reflection on peer feedback. In addition, interactive learning activities can help modify student behavior in ways that make students act more like instructors. For example, the in-class engagement activity of voting helped students participate in feedback exchange as long as the instructors did. The collaborative guiding question activity helped students recognize the value of Improve questions and devalue Share questions, as the instructor suggested was most effective. The collaborative reflection protocol helped students focus on using feedback to iterate their work, which was the instructor's goal for feedback exchange.

When implemented thoughtfully, in-class technology helps rather than distracts students.

When we first began this line of research, we found that nearly every existing peer feedback system had been designed for use outside of class. When talking with instructors, we found that they feared technology during class would distract their students from learning. However, this work shows that when in-class technology is introduced with scaffolding such as interactive learning activities and training in best practices and is used in conjunction with established learning goals, technology can support student learning and even decrease distraction. Students in Study 1 spent more time on task and less time on Facebook because the in-class technology they used provided relevant and interactive ways for them to engage in the course content. Other in-class systems, regardless of whether they relate to feedback, could use the principles of interactive learning activities and alignment with learning goals to be more effective and less distracting.

Classroom culture and team dynamics are essential elements of peer feedback exchange.

In every chapter of this dissertation, how students interact with each other on a social level impacted how they learned from peer feedback exchange. In Chapter 4, student behavior indicated their awareness of these social aspects of feedback exchange. Students appreciated the opportunity to be anonymous and often chose unidentifiable usernames. Students did not downvote other's comments often, and they drew encouraging "stickers" on the paper notecards for their peers. In Chapter 5, the ability to work with a team when writing questions influenced the type of questions written and therefore the quality of feedback received. In Chapter 6 and Chapter 7, students revealed that they often come to team reflection meetings expecting the reflection process to be a team building activity. When students use different reflection strategies in team meetings, it impacts not only team dynamics but also the next step students want to take on their project. At every phase of peer feedback exchange, student interaction impacts learning. This could be why interactive learning activities had such a strong, positive impact in the studies we conducted. Future work in this area could do more to identify the impact of team dynamics and design activities to more explicitly support team building.

Future Work

There are many avenues for future work that builds on this dissertation. I briefly explore some of those potential avenues here.

What Happens at a Different Scale?

One way to change the scale would be to increase the number of students in the course. This not only increases the number of presentations, but also increases the quantity of feedback each presenter must process. This raises questions like "Is there an upper bound for how much feedback is helpful?" and "What sort of system support is needed when feedback reaches a certain volume?". The larger scale might also require different classroom logistics. For example, if there are too many presentations to have each group present one after another, could a poster session format be effective for getting high-quality formative feedback on student work?

Another way to change the scale would be to increase the amount of time teams are working on the project. For example, say teams are working on a project for an entire year. Eventually, teams will need help deciding what *not* to iterate as well as suggestions for where to focus their efforts. This raises questions like "Will 'improve' questions still be the most helpful after 52 weeks?" and "What sort of system support is needed at different stages of a projects?". In addition, for longer projects, teams will also likely need to pay more attention to team dynamics, as those factors could make or break a long-term project. This raises questions like "How can system design influence or improve team dynamics in long-term projects?".

What Automated Support Could a System Provide?

There are so many ways to explore automated support in digital peer feedback systems. I will offer just three in this section.

Automated Quality Suggestions

We know that students have trouble writing high-quality feedback, even when they are given training ahead of time. What if the system could provide real-time suggestions to students in the moment they are writing their feedback? For example, the system could use natural language processing to detect if a student is writing a "descriptive" comment, then suggest that the student rephrase to make it more "actionable".

"Jigsaw" Feedback

In the jigsaw method, students split up the work, each team member taking responsibility for one task or concept. This can be applied to peer feedback. For example, in a game design class, each student has a particular expertise (art, sound design, game design, producer, etc.). What if the system used a matching process to ask the right person to answer the "right" feedback question based on their expertise? In addition using this distributed expertise to get higher-quality feedback, this might also help address emotional or social aspects of feedback exchange. The artist who gets chosen to help with that team's art-related question might think, "oh, they want MY feedback on this?" in a way that could foster intrinsic motivation to do a better job as a feedback provider.

Supporting Interactivity for Individual Projects

In this dissertation, we leveraged the fact that students were working on teams to create moments for interactive learning. However, students often complete individual projects. In these cases, where there is no group to interact with, what if we used the system to create these opportunities for interaction? One way to do this might be to use the system to create connections among individuals. For example, it might see that two students asked similar questions about their project, and encourage them to re-write their question together. Another way might be to use the system as the interactive agent. Perhaps the system includes an AI avatar that helps you reflect on the feedback you receive, for example.

What Happens in Other Disciplines?

In this dissertation, we explored feedback exchange in game design classrooms. One reason we chose game design is that this discipline has inherent support for feedback as a valuable learning activity because feedback exchange is part of the job description for professional game designers. This means that instructors see the value in feedback exchange and help express that value to students, either through direct instruction, modeling good feedback behavior, or having feedback skills as learning goals for the course. However, other project-based disciplines that could benefit from improved feedback might not have feedback already set up as such an important aspect of the discipline's culture. Future work should explore the challenges that arise in these settings and how system design might be able to influence student and instructor feelings towards feedback exchange.

9 References:

[1] Anderson, L.W. et al. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*.

[2] Andrade, H. and Du, Y. 2005. Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*. (2005).

[3] Andrade, H.G. 2005. Teaching with Rubrics: The Good, the Bad, and the Ugly. *College Teaching*. (2005).

[4] Beyer, H. and Holtzblatt, K. 1997. *Contextual Design: Defining Customer-Centered Systems*. Elsevier.

[5] Blau, I. and Caspi, A. 2009. What type of collaboration helps? Psychological ownership, perceived learning and outcome quality of collaboration using Google Docs. *Proceedings of the Chais conference on instructional technologies research* (2009).

[6] Bolen, J. 2009. *An investigation of limited professional development on teacher questioning and learner responses*. Walden University.

[7] Buabeng-Andoh, C. 2012. Factors influencing teachers' adoption and integration of information and communication technology into teaching: A review of the literature. *International Journal of Education and Development using Information and Communication Technology*. 8, 1 (2012), 136–155.

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

[9] Camerer, C. et al. 1989. The curse of knowledge in economic settings: an experimental analysis. *Journal of Political Economy*. 97 (1989), 1232–1254.

[10] Chen, N.-S. et al. 2009. Effects of high level prompts and peer assessment on online learners' reflection levels. *Computers & Education*. 52, 2 (2009), 283–291.

[11] Chester, A. and Gwynne, G. 2006. Online Teaching: Encouraging Collaboration through Anonymity. *Journal of Computer-Mediated Communication*. 4, 2 (Jun. 2006).

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

[13] 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.

[14] Cook, A. et al. 2019. How Guiding Questions Facilitate Feedback Exchange in Project-Based Learning. *ACM Conference on Human Factors in Computing Systems (CHI)* (2019).

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

[16] 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.

[17] Ertmer, P.A. et al. 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.

[18] Falchikov, N. 2001. *Learning Together: Peer Tutoring in Higher Education*. Psychology Press.

[19] Feedbackr: *feedbackr.io*. Accessed: 2015-09-10.

[20] Fies, C. and Marshall, J. 2006. Classroom response systems: A review of the literature. *Journal of Science Education and Technology*. (2006).

[21] 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.

[22] Fried, C.B. 2008. In-class laptop use and its effects on student learning. *Computers & Education*. 50, 3 (Apr. 2008), 906–914.
[23] Galloway, C.G. and Mickelson, N.I. 1973. Improving Teachers' Questions. *The Elementary School Journal*. 74, 3 (Dec. 1973), 145–148.
[24] Gibbs, G. and Simpson, C. 2004. Conditions under which assessment supports students' learning. *Learning and teaching in higher education*. (2004).
[25] Gielen, S. et al. 2010. Improving the effectiveness of peer feedback for learning. *Learning and Instruction*. 20, 4 (Aug. 2010), 304–315.
[26] Harboe, G. and Huang, E.M. 2015. Real-World Affinity Diagramming Practices. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (New York, New York, USA, 2015), 95–104.
[27] Hattie, J. and Timperley, H. 2007. The Power of Feedback. *Review of Educational Research*. 77, 1 (Mar. 2007), 81–112.
[28] Hestenes, D. et al. 1992. Force concept inventory. *The Physics Teacher*. 30, 3 (Mar. 1992), 141–158.
[29] Hinds, P. and Mortensen, M. 2005. Understanding Conflict in Geographically Distributed Teams. *Organization Science*. (2005).
[30] Howley, I. 2015. *Leveraging Educational Technology to Overcome Social Obstacles to Help Seeking*. Carnegie Mellon University.
[31] i>Clicker: *https://www1.iclicker.com/for-education/*.
[32] Kapsul: *http://www.kapsul.org/about*. Accessed: 2015-09-10.
[33] Kolb, D.A. 1984. *Experiential Learning*. Prentice-Hall.
[34] Kourilsky, M. 1993. Economic Education and a Generative Model of Mislearning and Recovery. *The Journal of Economic Education*. 24, 1 (Jan. 1993), 23–33.
[35] Kulkarni, C. et al. 2013. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction*. 20, 6 (Dec. 2013), 1–31.
[36] 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.
[37] Lave, J. and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
[38] Li, L. et al. 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.
[39] Liu, N.-F. and Carless, D. 2006. Peer feedback: the learning element of peer assessment. *Teaching in Higher education*. (2006).
[40] Liu, W. et al. 2018. ConsensUs. *ACM Transactions on Social Computing*. 1, 1 (Jan. 2018), 1–26.
[41] Ma, X. et al. 2015. Exiting the Design Studio. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15* (New York, New York, USA, Feb. 2015), 676–685.
[42] Mandala, M. et al. 2017. Comparison of Collective Team and Individual Student Peer Feedback on Design. *Volume 3: 19th International Conference on Advanced Vehicle Technologies; 14th International Conference on Design Education; 10th Frontiers in Biomedical Devices* (Aug. 2017), V003T04A017.
[43] 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.
[44] 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.

[45] Mumtaz, S. 2000. Factors affecting teachers' use of information and communications technology: a review of the literature. *Journal of Information Techology for Teacher Education*. 9, 3 (Oct. 2000), 319–342.
[46] 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).
[47] Oliveira, A.W. 2010. Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching*. 47, 4 (Apr. 2010), 422–453.
[48] Orsmond, P. et al. 2000. The Use of Student Derived Marking Criteria in Peer and Self-Assessment. *Assessment & Evaluation in Higher Education*. (2000).
[49] Piper, A.M. and Hollan, J.D. 2009. Tabletop displays for small group study. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (New York, New York, USA, Apr. 2009), 1227.
[50] Pol.is: *https://pol.is/2arcefpshi*. Accessed: 2015-09-10.
[51] Prince, M. 2004. Does active learning work? A review of the research. *Journal of engineering education*. (2004).
[52] 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.
[53] Ratto, M. et al. 2003. The ActiveClass Project: Experiments in Encouraging Classroom Participation. *Computer Support for Collaboartive Learning*. (2003).
[54] Reddy, Y.M. and Andrade, H. 2009. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*. (Aug. 2009).
[55] Sadler, D.R. 1989. Formative assessment and the design of instructional systems. *Instructional Science*. 18, 2 (Jun. 1989), 119–144.
[56] 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.
[57] Sellen, A.J. and Harper, R.H.R. 2003. *The Myth of the Paperless Office*. MIT Press.
[58] Shannon, A. et al. 2017. Better Organization or a Source of Distraction? Introducing Digital Peer Feedback to a Paper-Based Classroom. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)* (2017).
[59] Shannon, A. et al. 2016. PeerPresents: A Web-Based System for In-Class Peer Feedback during Student Presentations. *Designing Interactive Systems* (2016), 447–458.
[60] 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).
[61] Topping, K. 1998. Peer assessment between students in colleges and universities. *Review of educational Research*. (1998).
[62] Treem, J.W. and Leonardi, P.M. 2013. Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association. *Annals of the International Communication Association*. 36, 1 (Jan. 2013), 143–189.
[63] 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.
[64] Yen, Y.-C.G. et al. 2017. Listen to Others, Listen to Yourself: Combining Feedback Review and Reflection to Improve Iterative Design. *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C&C '17)* (2017), 158–170.
[65] 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.
[66] Zhu, H. et al. 2014. Reviewing versus doing. *Proceedings of the 17th ACM conference on Computer*

supported cooperative work & social computing - CSCW '14 (New York, New York, USA, Feb. 2014), 1445–1455.
[67] 1997. TurnItIn.