Increasing Quality and Involvement in Online Peer Feedback Exchange

SNEHA R. KRISHNA KUMARAN, DEANA C. MCDONAGH, and BRIAN P. BAILEY, University of Illinois at Urbana-Champaign, USA

Design instructors are integrating the use of online peer review platforms to keep pace with growing class sizes. However, these platforms typically prioritize randomized peer assignment strategies and show only the current solution to peers when writing feedback. This can result in low quality feedback in project-based design courses. We report on an experiment in which students (N=59) worked on twelve-week design projects and both wrote and received online feedback at four stages. The experiment tested a novel assignment strategy of peer mentorship, where peers were assigned to give feedback to all stages of the same project, and tested showing the context from the preceding design stage when composing feedback. The results showed that displaying the context from the preceding design stage led to feedback with higher perceived quality at the late design stages (but not at earlier stages) and feedback from mentors prompts longer responses from the feedback recipients. Our work contributes deeper empirical understanding of how assignment strategies and showing additional context affects peer feedback and provides practical guidelines for instructors to implement these methods in design courses.

CCS Concepts: • **Group and Organization Interfaces → Collaborative Computing**;

Additional Key Words and Phrases: Peer feedback; learning; design; peer mentorship; iteration

ACM Reference Format:
Sneha R. Krishna Kumaran, Deana C. McDonagh, and Brian P. Bailey. 2017. Increasing Quality and Involvement in Online Peer Feedback Exchange. *Proc. ACM Hum.-Comput. Interact.* 1, 1, Article 63 (November 2017), 18 pages.
https://doi.org/10.1145/3134698

1 INTRODUCTION

Peer review is a staple of design education. Through peer review, students with similar abilities assess and learn from each other's work [21]. Peer review has traditionally been conducted face-to-face in design courses [7]. However, as demand for design education increases, design instructors are increasingly integrating the use of online peer review platforms into their courses. Although this is scalable, prior studies have shown that online peer review can produce feedback that is narrow in scope and of limited quality [33, 34, 39].

To improve peer review, prior research has proposed platform features such as anonymity [12, 23], rubrics [4], and structuring the feedback assessment [10], and has tested how these features affect the feedback exchange. Researchers have also studied features aimed at increasing student interaction. For example, Talkabout [19] connects peers for topic discussions in online courses based on their geography to improve student engagement and learning. PeerStudio [20] dynamically

2573-0142/2017/11-ART63
https://doi.org/10.1145/3134698

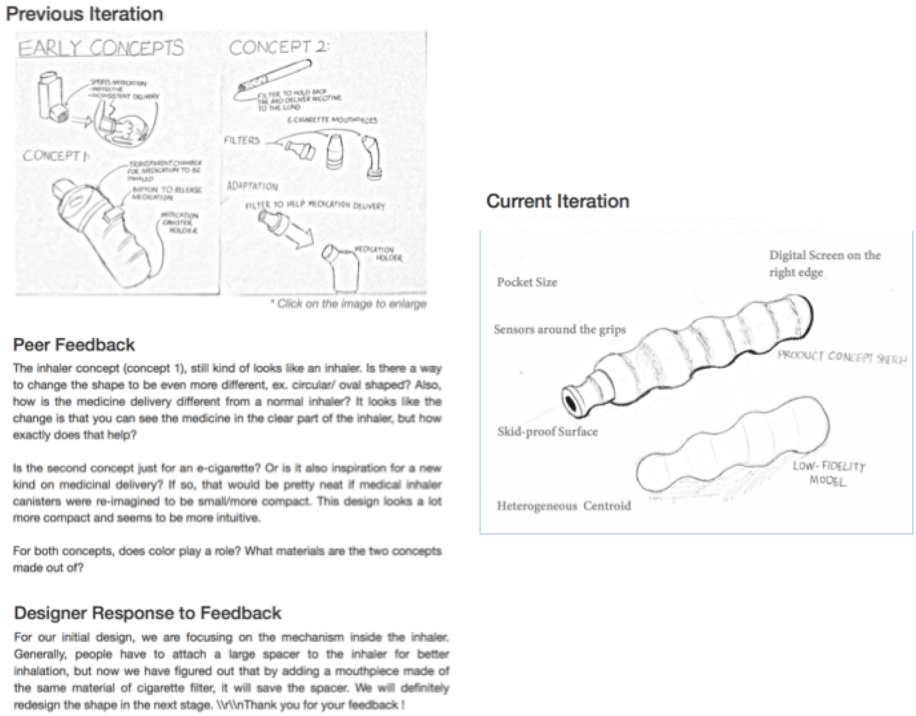

Figure 1: In the context condition, our platform displays the prototype(s) from the prior design stage, the peer feedback received on that prototype, and the designer's response to that feedback (context—right image). Students in this condition can review the context when writing feedback for the current prototype (left image). The current prototype and context were shown on separate pages in the actual implementation.

recruits providers for rapid peer feedback at a student's request, while PRAZE [26] allows providers and receivers to interact via forum-type feedback tools.

Our research contributes to this thread of prior work by challenging two assumptions common in many peer review platforms. First is the need for randomized assignment of classroom peers to submissions for review. Randomization can reduce potential biases and promote perceived fairness [15], yet it is impersonal and reduces the social interaction and ensuing connections often formed through face-to-face critique [16]. The second assumption is that a solution submission (e.g., a prototype for a design project) should be assessed independent of its prior solutions (e.g., the prior prototype for that project). This assumption is valid in some courses, but is less valid for project-based courses in which student solutions progress through multiple stages.

In this paper, we challenge these assumptions by proposing and empirically testing two novel features for a peer review platform in a project-based design course: peer mentorship and providing the context of a project's progression. For peer mentorship, a classroom peer is assigned to provide feedback to the same project at each design stage. This was contrasted with a control condition in which peers were randomly assigned to projects at each stage. For the context feature, the review platform presents the prototype as well as the prototype from the prior design stage, the feedback received at that stage, and the designer's response to that feedback (see Figure 1). In a control condition for this feature, only the current prototype was shown for the assessment.

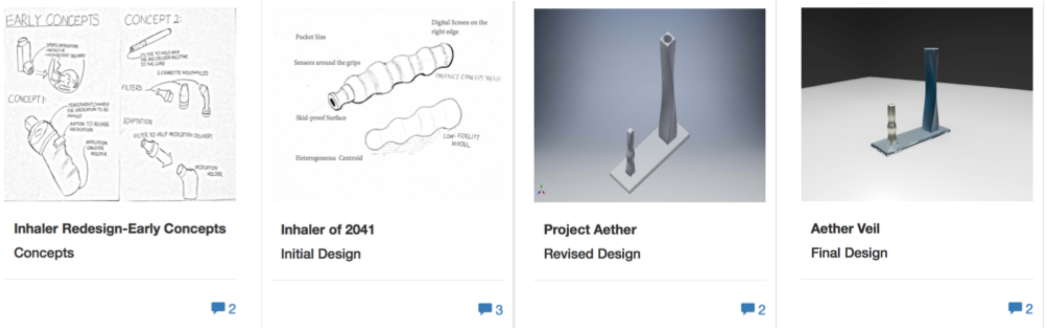

Figure 2: An example of a design project in the course studied. The team's initial concept was a redesign of an inhaler. After receiving peer feedback questioning the lack of futurism required in the design brief, the team changed the direction of the project to a concept that delivers medicine to patients through air circulation.

We hypothesize that when a feedback provider is assigned as a mentor and presented with the context of a prototype's progression, the provider will perceive a stronger social bond to the designer (and project), which, in turn, should have a positive impact on feedback composition. To test this hypothesis, we conducted a controlled experiment in a product design course. In the course, students (N=59) worked in teams (2 to 3 students) to develop a solution for a given product design brief (see Figures 1 and 2). The project required teams to submit prototypes and exchange feedback at four stages (concept, low-fidelity, medium-fidelity, and high-fidelity stages). Students in the course served as both designers and feedback providers. For the experiment, students were assigned either as a peer mentor for one project or randomly to many projects throughout the course. The students in each assignment strategy condition were further allocated to one of two context conditions: shown vs. not shown.

At each design stage, we measured the perceived quality and length of the feedback, categorized its discourse, and labeled its sentiment. We also had designers report the actions taken on the feedback received. A survey [5] was used to measure social bond from the perspective of the feedback provider after writing the feedback at each stage. We analyzed how the features affected social bond and how social bond mediated the relations between the features tested and the feedback measures. At the end of the course, surveys and interviews were conducted with students in the course to understand the benefits and weaknesses of the experimental conditions and the feedback activity.

Our study had two main findings. First, students in their role as designers wrote longer responses to the feedback provided by mentors. When assigned as mentors, students reported being more receptive to the feedback received compared to students who wrote feedback for randomly assigned projects. Second, our results show that the feedback exchanged at the late design stages was of higher perceived quality when the context was shown. Showing the context at earlier stages resulted in feedback that was perceived to be of lower quality, indicating the context was more of a distraction that an aid at these stages.

Our work makes three contributions to the CSCW community. First, our results contribute deeper empirical understanding of how peer mentorship and showing context of a prototype affects peer feedback exchange in a project-based design course. Second, we provide guidelines for design instructors to determine when to use these features (e.g., show context only at the later stages of a project). Third, our results have implications for the design of peer feedback platforms (e.g., offer

more flexible peer assignment strategies and allow students to decide what information to share with peer reviewers).

2 RELATED WORK

We describe how the features tested in our experiment draw from social bond theory, and how testing these features contributes to the base of knowledge for creating effective peer review platforms in learning environments.

2.1 Social Bond Theory

Social bond is a widely discussed and supported sociological theory of deviant behavior [1, 6]. Social bond theorizes that undesirable behavior occurs when the relation between an individual and community is lacking [11]. Though the theory was originally developed to explain criminal behavior, it has since been applied in a variety of settings including customer retention [32], information security [30], virtual communities [9], and education [5]. In learning environments, social bond consists of:

Attachment: Emotional closeness to peers and a community of learners.
Commitment: Rational calculation of the costs of undesirable social behavior for future goals.
Involvement: Perceived and actual investment in expected learning activities.
Belief: Acceptance of expected learning goals [11].

In face-to-face courses, social bond is typically developed through instructor-student and peer interactions in the classroom. However, as the use of educational technology begins to move these interactions online, peers may feel less attached to each other, less committed to the course goals, and less involved in the learning activity. This could explain why studies of online peer review have shown that the content often has limited scope and quality [33, 36, 39].

In our work, we use social bond to motivate two novel features for a peer feedback platform: peer mentorship in which a student is assigned to review the same project at each design stage and showing the context of the prototype at the prior design stage during feedback composition. A peer mentor may perceive stronger attachment to the assigned project because the designer would be perceived as being more dependent on the mentor's feedback. Likewise, investing the effort needed to review the context for a prototype may cause a provider to be more committed to the task of writing feedback for the project and to the course goals. Through increased social bond, these features could lead to higher quality feedback. The proposed features are not exhaustive, rather they represent a starting point for considering how feedback platforms can nurture the social bonds between peers interacting online.

2.2 Peer Feedback in the Classroom

Peer feedback asks students similar in ability to assess each other's in-progress work [21]. Peer feedback supports skill assessment and learning through exposure to different solution strategies [8, 29], and can lead to improved solution quality [14, 29]. To facilitate peer feedback, many design instructors have students present their prototypes in-class and facilitate a discussion. Though this approach can yield productive discussions, students may not fully participate due to social anxiety, evaluation apprehension, and the fear of voicing criticism in public [2, 17, 25, 27, 35]. This approach also does not scale well to large classes [34].

To overcome these limitations, many instructors are incorporating the use of online platforms to facilitate peer feedback exchange [34]. A goal of using these platforms is to maintain the learning benefits, while also leveraging features to further improve feedback quality. To meet this goal, researchers have tested novel features such as the use of rubrics [39], style guides [18], rapid

feedback [20], and anonymity [12]. Our work contributes to this body of work by testing a strategy of repeat peer assignment (mentorship) and showing the context of a project's progression during peer evaluation. These features are not exclusive and can be incorporated into existing peer feedback platforms.

Researchers have also investigated using external crowds to provide feedback in the classroom [13, 34, 37]. However, a prior study showed that peers provide higher quality feedback than external crowds for student design projects [34]. This suggests that peer feedback will continue to be a valued source of feedback in learning environments.

2.3 Technologies Supporting Peer Feedback

Among the results from multiple experiments, Hicks et. al. found that showing a simulated prototype and its process causes a provider to write more process-level feedback [10]. Our experiment extends this prior work by testing a similar experimental condition for authentic student projects and at multiple design stages, and by testing this condition with both random and repeat (mentorship) peer assignment strategies. PRAZE [26] implements a system in which a student responds to the feedback they received to mimic scientific peer review. In our experiment, we ask designers to respond to the feedback they received, but also test the benefit of feeding it forward to the next round of feedback for the project.

Researchers have also explored strategies for assigning peers to review each other's work. For example, Staubitz et al. [31] introduces a strategy that distributes reviews such that students who write more reviews also receive more reviews. Piech et al. [28] investigate a machine-learning model that can match peer graders to other students based on factors such as student engagement. Our work explores mentorship as another strategy for peer assignment. The rationale is to increase the perceived attachment to the assigned project and designer and belief that the designer will value the provider's feedback.

To assess the impact of the proposed features on feedback generation, we measured the perceived quality and length of the feedback, labeled its sentiment, and categorized its content using a taxonomy of critique discourse. Though not exhaustive, these measures are consistent with prior work in the CSCW community on peer feedback [4, 10, 18, 34, 38].

3 RESEARCH QUESTIONS

We address the following research questions:

- **RQ1.** How does peer mentorship and presenting context for a prototype affect the feedback written by a provider, e.g., its perceived quality, quantity, and discourse?
- **RQ2.** How does peer mentorship and presenting context for a prototype for peer feedback exchange affect the social bond experienced by the feedback providers?
- **RQ3.** How does social bond mediate the relationship between the features tested (mentorship and context) and the dependent measures relating to the feedback?
- **RQ4.** What do students perceive as the strengths and weaknesses of peer mentorship and presenting additional context of a prototype for writing feedback?

Answers to these questions will provide instructors with empirical knowledge and practical guidelines to further improve online peer feedback exchange. Answers will also have implications for the design of peer review platforms.

4 METHOD

We conducted a 2x2x3 experiment in a design course with two between-subject factors and one within-subject factor. The between-subject factors were Context (Shown vs. Not-shown) and

Assignment strategy (Mentor vs. Random). The within-subject factor was Stage (three project deliverables during the semester that required a submission). The experiment was approved by the IRB at our university.

4.1 Design Course and Projects

The experiment was conducted in an undergraduate human-centered product design course. The course has traditionally relied on the use of in-class critique for project assessment and feedback, but has recently integrated the use of online peer review to keep pace with increasing enrollment.

There were 59 students enrolled in the course. The course consisted of two smaller sections and the same instructor taught both sections. The two sections met at different times during the week and were balanced in terms of enrollment. A student was assigned to and attended only one section. All students consented for the data collected during this experiment to be used for research.

The main learning outcome for this course was a 12-week project worth 75% of the course grade. The duration of the course was sixteen weeks. Students were asked to design a futuristic product that would be useful in the year 2041 that solved a relevant societal problem. The course instructor randomly assigned students into groups of two or three. The course instructor evaluated all designs on futurism and usability. Students were expected to have a clear target user and scenario that their project was intended to address. Students generated a diverse set of design concepts, such as a novel medicine delivery device (see Figure 2), a cup that held fluids in a magnetic field, and a jacket that could be easily worn by the physically disabled.

The projects were structured as a design process. Each stage corresponded to a project deliverable. Students created a concept prototype (Stage I: Week 5), a low-fidelity prototype (Stage II: Week 8), a medium-fidelity prototype (Stage III: Week 11), and a high-fidelity prototype (Stage IV: Week 13—close to final model). At each of these stages, students presented their prototypes within their section of the course for five minutes and received short verbal feedback from the instructor and course assistant.

The use of online peer review was incorporated into each stage of the project. We adapted an existing online review platform (described in [38]) to collect and present the feedback at each stage. Participation in the feedback assignments accounted for 15% of the final course grade.

To enable anonymity between the designers and feedback providers and offer additional perspectives on a project, students exchanged feedback on the prototypes between the two sections. The platform did not show the identity of the designers or the feedback providers. In addition, students were asked not to discuss the prototypes they provided feedback on with their peers outside of class. The feedback exchange was anonymous because prior research shows that anonymity breeds more open and honest feedback and that designers typically perceive anonymous feedback to be more helpful [12, 23].

4.2 Peer Assignment Strategy

Peer assignment had two conditions: Mentor and Random. Students were randomly placed into one of these two conditions. Students in the Mentor condition were assigned to and gave feedback for the same team's project at all design stages. Students in the Random condition were assigned randomly to a team at each stage to mimic how most existing platforms assign deliverables to students for peer review.

Mentors were informed at the time of feedback composition that, "As a peer mentor, you will be able to see the progress of this project throughout the term. The feedback you provide will help designers determine how they are progressing from the previous checkpoint and provide another perspective." Randomly assigned providers were told, "You will provide feedback to this design for just this stage. The feedback you write will provide a new perspective to the designers of this

project." Students in the Context (shown) condition were additionally encouraged "to review the previous design, the feedback for the previous iteration, and the designer's response to the feedback when providing [...] feedback."

4.3 Feedback Content

Half of the students in each Peer Assignment condition were further placed randomly into a Context condition: Shown or Not Shown. The students in the Shown condition were shown the current prototype (as a single image with a description), the prototype(s) from the prior stage of the project, the feedback from the prior stage, and the designer's response to that feedback. See Figure 1. Students in the Not Shown condition saw only the current prototype as in existing peer feedback systems.

Because there is no context available at the first (concept) stage of the design process, we excluded data from the first stage from analysis. Instead, Stage I was used to introduce students to the platform and the procedures of the class.

4.4 Participants

The students enrolled in the course were undergraduate students in their third year of study. All but three students were pursuing a degree in industrial or graphic design. There were 33 females in the course (out of 59 total) and 18 students were from countries outside the U.S. In this study, students served as both designers and feedback providers.

In total, there were 16 students in the Random, Not-Shown condition; 14 students in the Random, Shown condition; 15 students in the Continuous, Not-Shown condition; and 14 students in the Continuous, Shown condition.

4.5 Procedure

At each of the four design stages of the project, teams uploaded visual representations of their prototype, along with a description of the project at that stage, to the peer feedback platform. Students had five days to write peer feedback to their assigned projects individually.

Teams uploaded their prototypes to the system at the beginning of the week. Students wrote the feedback later in the week and class time was allocated for writing it. Each student in the class individually (regardless of their team) wrote peer feedback to another team from the other section of the class. Students were asked to bring personal devices to class on these days. If students were not able to attend class or bring a device, they were allowed to give feedback remotely. At the end of the week, teams could access the feedback for their prototype through the platform. This ensured that all teams received feedback at the same time. Each team received at least two pieces of feedback on the current prototype: one piece of feedback from a student in the Mentor condition and at least one piece of feedback from a student in the Random condition. Once the team reviewed the feedback received, each team member was asked to individually rate the perceived quality of the feedback, indicate whether they would implement the feedback or not, and write a short response to the feedback they received. This response was shown to the students in the Context (Shown) condition during the next peer feedback session.

After providing feedback at each stage, students completed a survey instrument measuring social bond. Students were also given a final survey at the end of the term. In addition, eight students participated in interviews (not graded, with $10 remuneration) after the completion of the class.

4.6 Measures

We measured social bond, attributes of the feedback, actions taken on the feedback, and the designer's response. We also interviewed students from each condition.

Table 1: The taxonomy of critique discourse used for categorizing the content of the feedback. The taxonomy was adapted from [3].

Category	Definition	Examples
Judgment	Critics reacted to what they saw and rendered some assessment of its quality	"The locking system seems great and very protective of the space."
Process-Oriented	Critics made statements or asked questions about the student's design approach or process	"I feel that this could be thought outside the box a little more."
Brainstorming	Critics asked questions or made statements about future imagined possibilities for the design	"Maybe the spoon can morph into a different utensil."
Interpretation	Critics reacted to what they saw and tried to make sense of the concept or product	"I am unsure whether the glasses in the picture are your models or just for filler."
Recommendation	Critics gave specific advice about a particular aspect of design	"You may consider having the shape resemble that of a cup."
Investigation	Critics requested information (typically by questioning) about the design or the design process	"How does the device charge?"
Free Association	Critics made reactive, associative statements about the design	Not found in our data
Comparison	Critics contrasted the design or design process with something else	"This reminds me of the Microsoft HoloLens project."
Identity Invoking	Critics made statements or asked questions to suggest that students consider the larger picture of themselves as designers in a future professional community	"Think about how your users can benefit from your product."
Other	Any statement that did not fall into the above categories	"Good work." "Go Cubs."

4.6.1 Feedback Attributes, Actions Taken, and Responses. We measured the perceived usefulness and length of the feedback, labeled its sentiment, and categorized its content. These measures have been previously used for analyzing online design feedback [18, 34, 38].

Designers individually rated the quality of the feedback their team received. Ratings were performed on a 7-point Likert item (1=Low, 7=High), and were blind to condition. Feedback length served as a proxy for the effort made by the feedback provider. We felt the quantity of feedback written was a reasonable indicator of effort because the assignment did not impose a length requirement.

To categorize the content of the feedback and label its sentiment, we partitioned the feedback set into idea units. An idea unit represents a coherent unit of thought. The partitioning yielded a total of 958 idea units. On average, one piece of feedback contained 5.3 idea units (σ=1.99). Idea units were categorized using a taxonomy of critique discourse [3]. This taxonomy has nine categories: judgment, process-oriented, brainstorming, interpretation, recommendation, investigation, free association, comparison, and identity invoking. A category "other" was added to capture the idea units that could not be placed in the existing categories. Definitions and examples of each of the categories are shown in Table 1. Two coders with experience working with this taxonomy assigned the labels. Cohen's Kappa, a measure of inter-rater reliability, was 0.668 in a sample set of 75 randomly chosen idea units (5% of the data set), which indicates substantial agreement between the two coders [22]. The two coders then split the idea units and categorized them independently.

For sentiment, two coders labeled each idea unit as positive, critical, neutral, or indeterminate. The idea unit was labeled indeterminate if the statement clearly contained both positive and negative elements. Cohen's Kappa was 0.78 in the 75-unit sample set, which indicates substantial agreement between the two coders. The two coders then each labeled half of the idea units independently.

Designers also individually indicated the action they would take on each piece of feedback: implement it, consider it, or ignore it. Finally, designers individually wrote a response to the

Table 2: The survey given to students after each peer feedback assignment. This survey, adapted from [5], measured social bond. It consisted of Likert items on a scale from 1 (strongly disagree) to 7 (strongly agree). Items relating to involvement were rated from 1 (minimal) to 5 (maximum). I=Involvement, B=Belief, A=Attachment, C=Commitment.

Q1.	How much effort did you put into evaluating the design? (I)
Q2.	I believe the feedback I wrote will help the designers improve their designs. (B)
Q3.	I have the expertise needed to help the design become better. (B)
Q4.	My feedback will be valued by the designer. (B)
Q5.	I would like to review the next iteration of the design I reviewed. (A)
Q6.	I can be open and honest in my feedback. (A)
Q7.	I valued the design I reviewed. (A)
Q8.	I value the effort of the designers. (A)
Q9.	Learning to become a better designer is important to me. (C)
Q10.	Giving good feedback is important to me. (C)
Q11.	This activity helped me learn how to write better feedback. (C)
Q12.	I was given all the information needed to effectively review the design.

feedback that included further describing their prototypes and project goals, whether they would use the feedback and why, and what the provider could do to give better feedback next time.

4.6.2 Social Bond. Social bond consists of four attributes: Belief, Attachment, Commitment, and Involvement. These attributes were measured using a validated survey [5], but the questions were phrased to fit the context of our experiment. The questions asked in this survey are shown in Table 2.

To test the reliability of the survey, we measured the internal consistency of the questions using Chronbach's alpha (α). Participants answered four questions about Belief (α=0.800), four questions about Attachment (α=0.827), three questions about Commitment (α=0.840), and two or three questions about Involvement (condition dependent–α=0.843). All attributes had high internal validity ($\alpha > 0.8$).

5 RESULTS

In total, students gave and received 180 pieces of feedback. A mixed effects model was used to analyze the quantitative data. Unless otherwise noted, the model included three fixed effects (two between subjects: Context, Assignment and one within subject: Stage) and a random effect (feedback provider ID). Adding the feedback provider ID as a random effect accounts for the individual differences in the provided feedback over design stage. The analysis was performed using the lme4 package available in R.

The mixed effects model was used to analyze the perceived quality, feedback length, and response length. Chi-Squared tests were used to analyze the distributions of feedback categorizations, sentiment, and actions taken.

5.1 Perceived Quality (RQ1)

To analyze perceived quality, we used the mixed model with an additional random effect of designer ID to account for each designer rating multiple pieces of feedback. The perceived quality ratings are summarized in Table 3 and Figure 3.

An ANOVA showed an interaction effect between Context and Stage ($F(1, 116)=7.21$, $p=0.01$, $\eta^2=0.49$). A post hoc t-test showed that at the low-fidelity stage (stage 2), providers wrote feedback perceived to be of higher quality when not shown context (μ=5.7, SE=0.14) compared to when they were shown context (μ=5.1, SE=0.19; $t=2.51$, $p=0.013$). At the high-fidelity stage (stage 4), the opposite pattern was found. Providers wrote feedback perceived to be of higher quality when

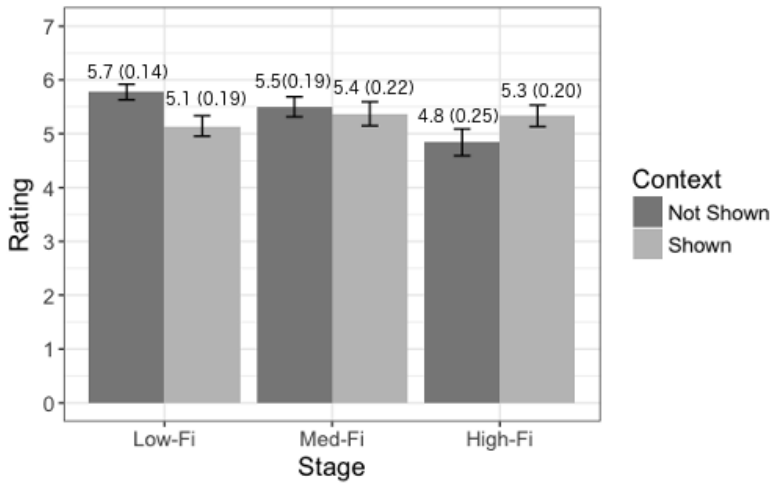

Figure 3: Ratings of perceived quality clustered by Context and Stage. The chart shows that showing context at the late design stage results in feedback of higher perceived quality than when context is not shown.

shown context (μ=5.3, SE=0.20) than when not shown context (μ=4.8, SE=0.25; t=2.057, p=0.04). See Figure 3.

This finding indicates that showing the context of a design is beneficial for feedback composition, but only for prototypes at the late stage. There was a main effect of Context, but this effect was due to the interaction. No other statistical effects were found.

5.2 Feedback Length (RQ1)

An ANOVA showed a main effect of Context (F(1,58)=3.5, p=0.06, η^2=0.17). Students who were not shown context (μ=115.5, SE=6.8) wrote more than students who were shown the context (μ=95.2, SE=6.4). A possible explanation is that students allocate an attention budget for the feedback task, and effort invested in evaluating the context of the prototype is subtracted from writing the feedback.

There was a main effect of Stage on feedback length (F(2,57)=9.2, p<0.01, η^2=0.44). Feedback length decreased at subsequent stages of a project as shown in Table 4, possibly indicating course fatigue. There were no other statistical effects found.

5.3 Responses to the Feedback (RQ1)

Designers wrote a response to the feedback they received blind to the experimental condition of the provider. To analyze the length of the response, we used the basic model replacing the random effect of feedback provider ID with the designer ID. An ANOVA found an interaction effect of Context and Assignment (F(1,58)=10.7, p<0.01, η^2=0.20). See Table 5. Students wrote more in response to the feedback from mentors who were not shown context (μ=66.9, SE=5.4) compared to students in the other conditions (μ=45.1, SE=1.8). Although mentors did not write feedback of higher quality, they wrote feedback that promotes a longer response. Results also showed that designers wrote shorter responses to providers who were shown context (μ=56.5, SE=3.1) than providers who were not (μ=43.4, SE=2.0; F(1,58)=10.3, p<0.01, η^2=0.20). As discussed earlier, providers who were shown context wrote less feedback, which may have prompted the designer to write shorter responses.

Table 3: Mean and (standard error) of the feedback quality ratings.

		Context			
		Not-Shown		Shown	
Assignment		Mentor	Random	Mentor	Random
Stage	Low-Fi	5.7 (0.24)	5.8 (0.16)	5.1 (0.28)	5.2 (0.26)
	Med-Fi	5.2 (0.28)	5.7 (0.24)	5.6 (0.31)	5.2 (0.31)
	High-Fi	5.0 (0.34)	4.6 (0.35)	5.4 (0.29)	5.2 (0.28)

Table 4: Mean word length of feedback and the designers' responses and (standard error) by Stage.

Stage	Feedback Length	Response Length
Low-Fi	126.3 (7.6)	59.6 (3.7)
Med-Fi	116.8 (8.1)	44.3 (2.9)
High-Fi	95.4 (5.1)	44.9 (2.7)

Table 5: The mean word length of designers' responses to feedback received and (standard error) by condition of the feedback provider.

	Context	
Assignment	Not-Shown	Shown
Random	48.2 (3.3)	45.9 (2.7)
Mentor	66.9 (5.4)	40.4 (3.1)

The analysis also shows a main effect of Stage on the length of the response ($F(2, 116)=12.9$, $p<0.01$, $\eta^2=0.25$) as shown in Table 4. Designers wrote less in response to the feedback at subsequent stages. Designers wrote the most at the low-fidelity stage (stage II; $\mu=59.6$, SE=3.7) compared to the medium-fidelity stage (stage III; $\mu=44.7$, SE=2.9) and the high-fidelity stage (stage IV; $\mu=44.8$, SE=2.7). Designers may have been reciprocating the effort perceived in the feedback received.

5.4 Content Categorization and Sentiment (RQ1)

Overall, judgment (37%), recommendation (20%), investigation (16%), and interpretation (11%) were the most referenced categories. The other categories were referenced much less: process-oriented (8%), brainstorming (3%), comparison (2%), identity invoking (<1%), and free association, was not commented on. A Chi-Squared test found no difference in the distribution of categories between the different Assignment and Context conditions or between the different stages.

The distribution of data between the categories corroborates prior work showing that novice feedback providers typically comment more on surface issues and less on conceptual and process issues [34, 38]. To address this imbalance, an instructor could provide rubrics to scaffold the generation of feedback that is most effective at each design stage [24, 39].

In the dataset, 35% of the idea units were labeled as critical, 22% as neutral, 36% as positive, and 7% as indeterminate. Our results showed that mentors wrote less critical feedback (31%) than those who were randomly assigned (39%; $\chi^2=9.31$, $p=0.002$). Perhaps mentors, being aware that they were going to review the same design at multiple stages, intentionally wrote feedback with a less negative tone. Context did not affect sentiment.

5.5 Actions Taken (RQ1)

For the actions taken, designers reported ignoring 12%, considering 58%, and implementing 30% of the feedback (percentages are for all designers overall). Neither Assignment nor Context affected the designer's reported actions on the feedback. However, our results show that designers who were mentors considered or implemented more feedback than students who were not mentors

Table 6: Final Survey–Questions and Results. This survey was completed at the completion of the course. Likert items were rated on a scale of 1 (strongly disagree) to 7 (strongly agree).

Q1.	I prefer to give feedback online rather than in person (Likert; μ=4.4, σ=1.9)
Q2.	I feel comfortable giving feedback online (Likert; μ=5.7, σ=1.5)
Q3.	I am interested in giving feedback for designs created by my peers (Likert; μ=6.0, σ=1.4)
Q4.	I learned how to write better feedback through the peer feedback activity (Likert; μ=4.9, σ=1.7)
Q5.	The feedback I received was helpful for improving my own design (Likert; μ=5.0, σ=1.5)
Q6.	I was interested in the designs that I reviewed (Likert; μ=5.4, σ=1.5)
Q7.	I would like to repeat this peer feedback activity in future design courses (Likert; μ=4.3, σ=2)
Q8.	What were the strengths of the feedback you received? (Open-ended)
Q9.	What were the weaknesses of the feedback you received? (Open-ended)
Q10.	What was the best thing about the peer feedback activity? (Open-ended)
Q11.	If there was anything I would change about the peer feedback activity, it would be: (Open-ended)

(χ^2=4.78, p=0.028). Non-mentors ignored twice (15%) the amount of feedback ignored by mentors (8%). An explanation for this finding is that mentors were more cognizant of the effort taken on the part of the feedback provider, which made them more receptive to the feedback they received. This finding is consistent with mentors providing more feedback with positive sentiment.

5.6 Effects On Social Bond (RQ2)

We conducted a separate ANOVA on each aspect of social bond (Belief, Commitment, Attachment, and Involvement). The ANOVA for Belief showed that students in the Mentor condition (μ=16.6, SE=0.52; scale ranges from 4 to 28) scored lower on the belief index than students in the Random condition (μ=17.1, SE=0.50; F(1,58)=3.48, p=0.065, η^2=0.34). One possibility is that because mentors were able to review the progression of the project, they could notice when their feedback was used or not. If their feedback was not used, they may not believe the designer valued their feedback. However, regardless of the belief the provider had, our data showed that the providers continued to write feedback of similar quality. This may have also impacted mentors to consider and implement more feedback as designers than the randomly assigned students.

The experimental factors did not have a statistical effect on Attachment (μ=23.6, SE=0.34; scale ranges from 4 to 28), Commitment (μ=18.7, SE=0.23; scale ranges from 3 to 21), or perceived Involvement (μ=8.6, SE=0.10; scale ranges 3 to 10). The high values on the measure of social bond (the means were at the upper end of the scales) indicates that the social bond may have been facilitated more by the in-class activities organized by the instructor compared to the online feedback exchange. No other significant effects were found.

5.7 Social Bond as a Mediator (RQ3)

For a mediation analysis, the independent variable must first be shown to correlate with the mediator variable. A mixed effects logistic regression with three fixed effects (Assignment, Context, and Stage) and one random effect (feedback provider ID) indicated that neither Assignment (b=-6.3, t=-1.0, p=0.34) nor Context (b=-3.3, t=-0.48, p=0.63) correlated with social bond. This result indicates that social bond did not mediate the effect of Assignment and Context on perceived quality or on the length of the feedback.

5.8 Perceived Strengths and Weaknesses (RQ4)

We report results of the final survey (Table 5; n=55) and interviews (n=8) conducted after the end of the course. In the final survey, students were asked four open-ended questions about the strengths and weaknesses of the feedback they received, their perceptions of the feedback activity, and any suggestions they had for improving the peer feedback activity. They were also asked seven Likert

item questions about their perceptions for giving feedback. The interviews further inquired about these issues and the student's experience in the assigned condition. We use "N=<value>" to indicate the number of responses citing a given theme.

5.8.1 Perceptions of the Peer Assignment Strategy. Students in both the mentor and random conditions commented on how they enjoyed the structure of the peer feedback assignment during the interviews and in the final surveys. Mentors appreciated that they could make a significant contribution to one project. A mentor for the mechanical baby caretaker (the Baby Pod) said,

> "I loved how I could see the project grow from start to finish. I could also see what parts of my feedback they used. I felt good about having a part in how well they did in the course." (P14, Mentor, Context Shown, Interview)

Another mentor, for a project that changed from a redesign of an inhaler to a multi-use laser pen, stated,

> "I felt a sense of responsibility towards the project I was mentoring. I put a lot of effort into my feedback because I knew they needed to make the product more futuristic." (P11, Mentor, Context Not Shown, Interview)

Students in the Random condition expressed wanting to have been assigned to one project for the course (N=15):

> "It would have been nice to see some of the designs progress each step of the way instead of only seeing them once in the term." (P20, Random, Context Shown, Survey)

While students felt that being a mentor for a project reduced cognitive load for giving a helpful critique, they still wanted to see other people's work:

> "I had one project for the entire duration and I got to see its changes and progress; however, I personally prefer [in-class] critiques because we can see everyone's designs and [...] learn from the critiques even if it doesn't apply to us." (P45, Mentor, Context Not Shown, Survey)

Though students mostly had a positive perception of being a mentor, there were weaknesses. One mentor, who wrote feedback for a team creating a language translator, did comment that she sometimes felt ignored by the designer:

> "I told them they needed to rethink the problem, because I didn't think it was persuasive enough, but they never addressed it. It was very frustrating." (P52, Mentor, Context Shown, Interview)

In the random condition, some students commented on how they enjoyed reviewing a variety of projects. For example, one student said:

> "I couldn't simply be a peer mentor. I would get bored seeing the same project again and again. I would want the option to see other designs too." (P39, Random, Context Shown, Interview).

A compromise could be to assign a student as a mentor to two or more projects. This would allow providers to interact with a more diverse set of projects and the designers would receive more feedback from mentors.

5.8.2 Perceptions of Context. Students in the context (shown) condition enjoyed seeing the previous prototype and felt it helped them write better feedback. One student said:

> "I used the previous design to figure out what direction the design was going and gave feedback accordingly." (P39, Random, Context shown, Interview)

The same student requested that we provide the entire history of the prototype starting from Stage 1. Another student (P14, Mentor, Context Shown) also requested more information about the history. Students also requested to be able to showcase their work in more than one image (a limitation of many existing feedback platforms).

From the perspective of a designer, students wished they had a way to communicate to providers what aspects of the prototype had already been commented on. In P20's words:

> "Some of the feedback was redundant because it was new people reviewing [the] designs." (P20, Random, Context Shown, Survey)

However, some feedback providers did not fully evaluate the context shown to them. P58 said,

> "I don't really remember [the feedback on the previous design or the responses]. I only used the prior design." (P58, Random, Context Shown, Interview)

5.8.3 Perceptions of the Feedback Activity. We comment on the perceptions of the feedback activity because the students in this course had not previously used an online platform for peer feedback exchange.

Students generally enjoyed the online feedback activity. Of the 55 students who responded to the final survey, only 18 students (evenly distributed by Assignment and Context conditions) indicated that they would not repeat this activity in the future. The remaining 37 students either wished to repeat the activity (16) or had no preference (21).

Many students (N=21) mentioned that they enjoyed the anonymity associated with the feedback activity. From the perspective of the role of feedback provider, students found that anonymity allowed them to be more open and honest. However, some designers perceived feedback to be extremely harsh. In the words of P46,

> "There was one person who was not very nice. Maybe it was out of their frustration of not understanding the design or because they could hide behind the keyboard and say whatever they want." (P46, Mentor, Context Not Shown, Survey)

Feedback providers also wished to see the designer's response to their feedback. To this end, students wanted a mechanism for discussion between the feedback provider and the designer.

Students also enjoyed getting feedback from a variety of people (N=19). They felt that getting feedback online increased participation of students who normally did not speak up in class. One non-native English speaker noted:

> "I don't normally participate in in-class feedback sessions because I do not think my English is very good. With this online platform, I could think through the design and slowly translate my thoughts." (P29, Random, Context Not Shown, Interview)

Moving critique online may provide a way for students with English as a second language to contribute more to the class critique.

6 DISCUSSION AND FUTURE WORK

A goal of this study was to test how being assigned as a peer mentor and how showing the context of a prototype's progress affects online peer feedback exchange.

Our results show that feedback providers who were assigned to be mentors wrote feedback that contained more idea units with positive sentiment, that prompted a longer response from the designer, and was of similar perceived quality compared to the feedback written by students in the random condition. We also found that students assigned as peer mentors were more receptive (i.e. reported ignoring less) of the feedback they received on their own projects compared to students who were randomly assigned to projects. From the final surveys, about half the students (15 of 29) in the random condition self-reported they would have preferred to be assigned as a peer mentor.

Providers who were shown context for the high-fidelity prototypes (stage 4) wrote feedback perceived to be of higher perceived quality than providers who were not shown context for the same prototypes. From the final surveys, students reported wanting to both review (as the provider) and reveal (as the designer) more of a project's progression for feedback exchange. When writing feedback, students reported wanting to view the previous prototypes, a list of the changes the designer made from the prior stage, and any changes the designer was considering at the current stage.

However, for low-fidelity prototypes (stage 2), providers wrote feedback of lower perceived quality when shown context compared to when they were not shown context. An explanation could be that some of the designers changed the conceptual direction of the project and the context from the prior stage was less relevant for the prototype at the current stage. This pattern of results indicates that showing the context of a prototype's progression should be preferred at the late design stages. Platforms could also give designers control over which context to make available to providers.

Our results also surfaced limitations of being assigned as a mentor. Mentors did not write feedback of higher perceived quality than non-mentors. Some mentors observed that the designers did not address the feedback given at the previous stage, or were assigned to projects that did not fit their interests, and were less motivated to write feedback. Mentorship also negatively influenced a provider's belief that the designers will use their feedback. This may be the result of the mentor recognizing instances where the designer did not apply his or her feedback from the prior stage. It may also have been due to the rigid structure of the feedback exchange in the course. Feedback was only provided and responded to once at each stage rather than unfolding as an open dialog.

Analysis of the feedback content showed that students referenced similar categories of feedback regardless of the experimental condition. The distribution of the ideas units across the categories corroborates findings from prior work showing that novice providers frequently recommend surface improvements and write judgments of a prototype, while infrequently commenting on the conceptual aspects of the prototype [34, 38]. In a course, this suggests that instructors should consider using rubrics to direct student attention to the most appropriate aspects of a project at each stage. It also may be useful for designers to request specific types of feedback and pose questions to the providers.

The findings of our experiment extend the results reported by Hicks et al. [10] by showing that providing context to a design during feedback composition increases the perceived quality of the feedback at the late design stage. Though context helps at the late stage, it was not effective at the earlier stages. The context may have been less effective at the early stages because some designers were changing the prototypes significantly because they were exploring different conceptual directions. Further, our results showed that providing context results in feedback that is shorter (about 20 words) than when context is not shown. This may be because the time required to interpret the context reduced the effort that students were willing to invest in writing the feedback.

The analysis of our data did not show a mediation effect of social bond between the proposed features and the feedback measures. The effects of the experimental features on social bond may have been masked by the effects of the in-class activities facilitated by the instructor. The implementation of platform features to nurture social bond may show a stronger effect in learning environments where more interactions occur online. However, our results do provide evidence that mentorship and providing context significantly benefit feedback quality and receptivity even in the presence of high social bond.

The results of our experiment are most generalizable to project-based courses that incorporate the use of online peer feedback exchange. Future work is needed to test the generalizability of our results to learning environments in which all the interactions occur online (e.g., MOOCS). Peer

mentorship and providing context of a prototype may show a larger benefit in this environment because students would not have the regular in-person engagement with the course staff and a cohort of peers as in our study. However, a future experiment is needed to test this claim.

To implement peer mentorship in project-based courses, instructors can create or adapt a platform similar to the one used in our experiment or implement feedback exchange in the online discussion forums available in many instructional technologies. Ideally the instructor can configure the forums for anonymous exchange and limit access to only the designer and the assigned mentors. As this forum also acts as a record of prior comments and design stages, feedback providers can also review a project's progression if needed. Instructors should also consider assigning mentors to two or more projects so students can contribute to the progression of a project, while also benefiting from the exposure to different projects and solution approaches.

The implementation of context in our experiment included the prototype at the prior stage, feedback for that prototype, and the designer's response to that feedback. Future work could study how providing different amounts of context and for additional prior stages of the project affects the feedback. Additionally, it would be interesting to allow designers to choose what context to show and learn from these choices. Another interesting direction would be to test how different visual representations of the context affect feedback exchange.

Beyond the issues already discussed, another direction for future work is to explore additional strategies for peer assignment beyond mentorship. For example, strategies could include assigning projects to feedback providers based on the topic of the project, student interest, or course performance. Another direction would be to test the effects of implementing a mentorship model for acquiring feedback from crowdsourcing and social media platforms.

7 LIMITATIONS

One limitation of this experiment is that we conducted it in only one university design course. Future work is needed to test the generalizability of our findings to other types of design courses and in different university cultures. Second, in this study, we had students write open-ended feedback to mimic the open-ended nature of in-person critique. However this approach did not take advantage of prior research that has shown that scaffolding the feedback exchange (e.g., by using rubrics) can improve the quality of feedback from less experienced designers [34, 37]. Third, while the experiment implemented anonymous feedback exchange, there is no guarantee that students maintained this anonymity outside of the course. Finally, we only measured social bond from the perspective of the providers because our focus was on improving feedback quality. Future work is needed to test if the designer perceives a stronger bond to providers assigned as peer mentors.

8 CONCLUSION

Design educators are increasingly using online peer review in their courses. This is generating tremendous research interest in how to promote quality feedback exchange online. This paper contributes empirical knowledge of how a strategy of repeat peer assignment (i.e., being a peer mentor) and presenting additional context for a prototype (e.g., the prototype at the prior stage and its feedback) affect online peer feedback exchange. Our findings show that being assigned as a peer mentor makes students more receptive to the feedback received on their own projects and that designers write more in response to feedback from mentors. Students who were assigned as a peer mentor enjoyed their role, while many students who were assigned randomly to projects reported they would have preferred to have been assigned as mentors. Our findings also show that providing additional context for a prototype is most helpful during the late (refinement) stages of a project. Our contributions add to the literature for features that can be implemented to improve

online peer review and provide instructors with practical guidelines for when to leverage these features in project-based courses.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation awards CMMI 14-62693 and IIS 15-30818.

REFERENCES

[1] Ronald L. Akers and Christine Sharon Sellers. 2009. *Criminological theories : introduction, evaluation, and application.* Oxford University Press, New York.

[2] Louise Arnold, Carolyn K. Shue, Barbara Kritt, Shiphra Ginsburg, and David T. Stern. 2005. Medical students' views on peer assessment of professionalism. *Journal of General Internal Medicine* 20, 9 (2005), 819–824. DOI: http://dx.doi.org/10.1111/j.1525-1497.2005.0162.x

[3] Deanna P. Dannels and Kelly Norris Martin. 2008. Critiquing Critiques. *Journal of Business and Technical Communication* 22, 2 (2008), 135–159. DOI: http://dx.doi.org/10.1177/1050651907311923

[4] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the Crowd Yields Better Work. In *ACM Conference on Computer Supported Cooperative Work*. 1013–1022. DOI: http://dx.doi.org/10.1145/2145204.2145355

[5] Keith K. Durkin, Timothy W. Wolfe, and Gregory Clark. 1999. Social bond theory and binge drinking among college students: A multivariate analysis. *College Student Journal* 33, 3 (1999). DOI: http://dx.doi.org/10.1080/027321790518681

[6] LeGrande Gardner and Donald J. Shoemaker. 1989. Social Bonding and Delinquency. *Sociological Quarterly* 30, 3 (1989), 481–500. DOI: http://dx.doi.org/10.1111/j.1533-8525.1989.tb01532.x

[7] Ilya M. Goldin, Kevin D Ashley, and Christian D. Schunn. 2012. Redesigning Educational Peer Review Interactions Using Computer Tools: An Introduction. *Journal of Writing Research* 4, 2 (2012), 111–119. DOI: http://dx.doi.org/10.17239/jowr-2012.04.02.1

[8] John Hattie and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77, 1 (2007), 81–112. DOI: http://dx.doi.org/10.3102/003465430298487

[9] F. Henri and B. Pudelko. 2003. Understanding and analysing activity and learning in virtual communities. *Journal of Computer Assisted Learning* 19, 4 (2003), 474–487. DOI: http://dx.doi.org/10.1046/j.0266-4909.2003.00051.x

[10] Catherine M. Hicks, Vineet Pandey, C. Ailie Fraser, and Scott Klemmer. 2016. Framing Feedback: Choosing Review Environment Features That Support High Quality Peer Assessment. In *ACM Conference on Human Factors in Computing Systems*. 458–469. DOI: http://dx.doi.org/10.1145/2858036.2858195

[11] Travis Hirschi. 1969. *Causes of delinquency.* University of California Press, Berkeley.

[12] Julie Hui, Amos Glenn, Rachel Jue, Elizabeth Gerber, and Steven Dow. 2015. Using anonymity and communal efforts to improve quality of crowdsourced feedback. In *Third AAAI Conference on Human Computation and Crowdsourcing*. 72–82.

[13] Julie S. Hui, Elizabeth M. Gerber, and Steven P. Dow. 2014. Crowd-based Design Activities: Helping Students Connect with Users Online. In *ACM Conference on Designing Interactive Systems*. 875–884. DOI: http://dx.doi.org/10.1145/2598510.2598538

[14] Christopher Hundhausen, Anukrati Agrawal, Dana Fairbrother, and Michael Trevisan. 2009. Integrating Pedagogical Code Reviews into a CS 1 Course: An Empirical Study. In *ACM Technical Symposium on Computer Science Education*. 291–295. DOI: http://dx.doi.org/10.1145/1508865.1508972

[15] Farnaz Jahanbakhsh, Wai-Tat Fu, Karrie Karahalios, Darko Marinov, and Brian Bailey. 2017. You Want Me to Work with Who? Stakeholder Perceptions of Automated Team Formation in Project-based Courses. In *ACM Conference on Human Factors in Computing System*. 3201–3212. DOI: http://dx.doi.org/10.1145/3025453.3026011

[16] Scott D. Johnson, Steven R. Aragon, Najmuddin Shaik, and Nilda Palma-Rivas. 2000. Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *J. Interact. Learn. Res.* 11, 1 (2000), 29–49.

[17] David A. Karp and William C. Yoels. 1976. The college classroom: Some observations on the meanings of student participation. *Sociology Social Research* 60, 4 (1976), 421–439.

[18] Markus Krause, Thomas Garncarz, Steven P Dow, Brian P Bailey, and Elizabeth Gerber. Critique Style Guide: Improving Crowdsourced Design Feedback with a Natural Language Model. In *ACM Conference on Human Factors in Computing Systems*. 4627–4639. DOI: http://dx.doi.org/10.1145/3025453.3025883

[19] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott R. Klemmer. 2015b. Talkabout: Making Distance Matter with Small Groups in Massive Classes. In *ACM Conference on Computer Supported Cooperative Work*. 1116–1128. DOI: http://dx.doi.org/10.1145/2675133.2675166

[20] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and self assessment in massive online classes. *ACM Trans. Comput.-Hum. Interact.* 20, 6 (2013), 1–31. DOI: http://dx.doi.org/10.1145/2505057

[21] Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. 2015a. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *ACM Conference on Learning @ Scale*. 75–84. DOI: http://dx.doi.org/10.1145/2724660.2724670

[22] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. DOI: http://dx.doi.org/10.2307/2529310

[23] Ruiling Lu and Linda Bol. 2007. A Comparison of Anonymous Versus Identifiable e-Peer Review on College Student Writing Performance and the Extent of Critical Feedback. *Journal of Interactive Online Learning* 6, 2 (2007), 100–115.

[24] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *ACM Conference on Computer Supported Cooperative Work*. 473–485. DOI: http://dx.doi.org/10.1145/2675133.2675283

[25] James C. McCroskey and John A. Daly. 1984. *Avoiding communication : shyness, reticence, and communication apprehension / John A. Daly, James C. McCroskey, editors.* Sage Publications, Beverly Hills. http://www.loc.gov/catdir/enhancements/fy0658/83027242-t.html

[26] Raoul A Mulder and Jon M Pearce. 2007. PRAZE: Innovating teaching through online peer review. In *Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*.

[27] Tracey Papinczak, Louise Young, and Michele Groves. 2007. Peer Assessment in Problem-Based Learning: A Qualitative Study. *Advances in Health Sciences Education* 12, 2 (2007), 169–186. DOI: http://dx.doi.org/10.1007/s10459-005-5046-6

[28] Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, and Daphne Koller. 2013. Tuned Models of Peer Assessment in MOOCs. In *International Conference on Educational Data Mining*. DOI: http://dx.doi.org/10.1080/02680513.2016.1188690

[29] D. Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18, 2 (1989), 119–144. DOI: http://dx.doi.org/10.1007/bf00117714

[30] Nader Sohrabi Safa, Rossouw Von Solms, and Steven Furnell. 2016. Information security policy compliance model in organizations. *Comput. Secur.* 56, C (2016), 70–82. DOI: http://dx.doi.org/10.1016/j.cose.2015.10.006

[31] Thomas Staubitz, Dominic Petrick, Matthias Bauer, Jan Renz, and Christoph Meinel. 2016. Improving the Peer Assessment Experience on MOOC Platforms. In *ACM Conference on Learning @ Scale*. 389–398. DOI: http://dx.doi.org/10.1145/2876034.2876043

[32] Karin A. Venetis and Pervez N. Ghauri. 2004. Service quality and customer retention: building longâĂŘterm relationships. *European Journal of Marketing* 38 (2004), 1577–1598. DOI: http://dx.doi.org/doi:10.1108/03090560410560254

[33] Mark E. Walvoord, Marielle H. Hoefnagels, Douglas D. Gaffin, Matthew M. Chumchal, and David A. Long. 2008. An analysis of calibrated peer review (CPR) in a science lecture classroom. *Journal of College Science Teaching* 37, 4 (2008), 66–73.

[34] Helen Wauck, Yu-Chun (Grace) Yen, Wai-Tat Fu, Elizabeth Gerber, Steven P. Dow, and Brian P. Bailey. 2017. From in the Class or in the Wild?: Peers Provide Better Design Feedback Than External Crowds. In *ACM Conference on Human Factors in Computing Systems*. 5580–5591. DOI: http://dx.doi.org/10.1145/3025453.3025477

[35] Robert R. Weaver and Jiang Qi. 2005. Classroom Organization and Participation: College Students' Perceptions. *The Journal of Higher Education* 76, 5 (2005), 570–601. DOI: http://dx.doi.org/10.1080/00221546.2005.11772299

[36] Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-experts. In *ACM Conference on Computer Supported Cooperative Work*. 1433–1444. DOI: http://dx.doi.org/10.1145/2531602.2531604

[37] Anbang Xu, Huaming Rao, Steven P. Dow, and Brian P. Bailey. 2015. A Classroom Study of Using Crowd Feedback in the Iterative Design Process. In *ACM Conference on Computer Supported Cooperative Work*. 1637–1648. DOI: http://dx.doi.org/10.1145/2675133.2675140

[38] Yu-Chun (Grace) Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 2016. Social Network, Web Forum, or Task Market?: Comparing Different Crowd Genres for Design Feedback Exchange. In *ACM Conference on Designing Interactive Systems*. 773–784. DOI: http://dx.doi.org/10.1145/2901790.2901820

[39] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *ACM Conference on Computer-Supported Cooperative Work*. 1005–1017. DOI: http://dx.doi.org/10.1145/2818048.2819953

Received May 2017; revised June 2017; accepted November 2017