From in the Class or in the Wild? Peers Provide Better Design Feedback Than External Crowds

Helen Wauck[1], Yu-Chun Yen[1], Wai-Tat Fu[1], Elizabeth Gerber[2], Steven P. Dow[3], Brian P. Bailey[1]

[1]University of Illinois at Urbana
{wauck2, yyen4, wfu, bpbailey}@illinois.edu

[2]Northwestern University
egerber@northwestern.edu

[3]University of California at San Diego
spdow@ucsd.edu

ABSTRACT
As demand for design education increases, instructors are struggling to provide timely, personalized feedback for student projects. Gathering feedback from classroom peers and external crowds offer scalable approaches, but there is little evidence of how they compare. We report on a study in which students (n=127) created early- and late-stage prototypes as part of nine-week projects. At each stage, students received feedback from peers and external crowds: their own social networks, online communities, and a task market. We measured the quality, quantity and valence of the feedback and the actions taken on it, and categorized its content using a taxonomy of critique discourse. The study found that peers produced feedback that was of higher perceived quality, acted upon more, and longer compared to the crowds. However, crowd feedback was found to be a viable supplement to peer feedback and students preferred it for projects targeting specialized audiences. Feedback from all sources spanned only a subset of the critique categories. Instructors may fill this gap by further scaffolding feedback generation. The study contributes insights for how to best utilize different feedback sources in project-based courses.

Author Keywords
Feedback; design methods; learning; crowdsourcing.

ACM Classification Keywords
H.5.3 [Information Interface and Presentation]: Group and Organization Interfaces--Collaborative computing.

INTRODUCTION
Formative feedback is a critical aspect of project-based design courses because it helps students assess and improve their in-progress work [11]. Without it, students learn less effectively and the quality of their project solutions suffers [14, 35]. However, instructors struggle to deliver timely, personalized feedback due to increasing student demand for design courses coupled with significant long-term resource constraints faced by many public institutions [2].

CHI 2017, May 06-11, 2017, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/3025453.3025477

Researchers have explored two approaches for scaling feedback generation in project-based design courses. The first engages classroom peers in feedback exchange [19]. Peer feedback has learning benefits for the recipient and the provider [8, 28], and there is a shared context to ground the exchange [6]. However, peer feedback can be narrow since students are learning the same course material and typically share similar demographics [11]. Peer feedback can also be affected by friendship and competition between the students [36], and the increased workload it imposes on them [22]. These issues could affect the quality of the feedback and students' willingness to act on it. Further, students depend on instructors to orchestrate the exchange.

The second approach engages online crowds external to the classroom. Students receive feedback from people in social networks, online communities, and task markets. External crowds may include more authentic, specialized audiences that would be hard to access in class. This might increase students' perceived value of and willingness to act on the feedback. Students can also gather feedback from external crowds as often as their time and resources allow. Planning for feedback aids self-regulation skills, which is critical for career readiness [43]. However, feedback from such crowds can be noisy due to the lack of a shared context [5], among many factors, and requires social or financial capital [41].

Despite the availability of both approaches in project-based courses, there is little knowledge of how they compare. Do students perceive differences in quality between feedback received from classroom peers and external crowds? Do they act on the feedback differently? Do these sources yield different categories of feedback that would be more or less suitable at different design stages? Investigating these questions will provide deeper empirical understanding of the relation between the feedback source and output, and help instructors devise scalable feedback strategies to provide the most help at different stages of student projects.

We conducted a study in which students (n=127) completed user interface design projects of their choice over nine weeks. Early- and late-stage prototypes were created during the projects. Students received feedback at each stage from three to six classroom peers and three crowds: their own social networks, online communities, and a task market. Students—blind to the source—rated the perceived quality and selected the action taken on each feedback response. We measured the number of responses, length, and valence of the feedback, and categorized the content using a critique taxonomy [4]. We also surveyed and interviewed students to gauge their perceptions of the feedback sources.

Our study has three main findings. First, classroom peers produced feedback that was of higher perceived quality, was acted upon more, and was longer compared to the feedback from the external crowds. This is likely due to the multiple incentives peers have for giving feedback: course grades, social ties in the classroom, and interest in the topic since this was an elective course.

Second, results show that feedback from external crowds is a viable supplement to peer feedback. This is supported by the findings that students perceived the crowd feedback to be of comparable quality (μ=4.1 vs. μ=5.0 for peers; 7-pt scale), students did not report feeling uncomfortable sharing their work with online crowds, and they preferred external crowds for accessing specialized audiences. Instructors can thus have increased confidence when appropriating the use of external crowds for design feedback in the classroom.

Finally, most of the feedback produced across all sources focused on recommending improvements, assessing design quality, and sense-making, while missing other categories of critique discourse (e.g., brainstorming possibilities and offering comparisons to existing solutions). Instructors should be mindful of this gap, especially at the early design stage, when the content from these missing categories may be most beneficial. Instructors can fill this content gap by further scaffolding the feedback generation process.

Our work makes three contributions to the CHI community. First, we offer deeper empirical knowledge of the relation between the source of the feedback provider and the quality, quantity, and content of their feedback. Second, our results offer guidance as to how instructors can better orchestrate feedback from different sources to benefit learning and improve project quality (e.g., organize peer feedback at project milestones and leverage external crowds for feedback in-between the milestones or to access specialized user audiences). These results can also inform how professional HCI designers consider the tradeoffs in obtaining feedback from colleagues or external crowds. Third, our study provides implications for improving the effectiveness of crowd platforms for feedback exchange.

RELATED WORK

We discuss tradeoffs of receiving design feedback from instructors, end users, classroom peers, and external crowds in large project-based courses. We highlight how the use of external crowds and classroom peers can address issues of scale and offer learning benefits. We also situate our work in context of prior HCI-related studies of these approaches.

Instructor and End User Feedback

Historically, instructor feedback has been considered an integral component of design education [37]. Instructors question the intent and effectiveness of the students' design work (sketches, renderings, 3D models); and students use the feedback to inform next steps [24]. Instructors can also draw from the literature to tailor the feedback to improve its learning benefit [31, 32]. However, it is difficult for instructors to provide personalized feedback to dozens of student projects in parallel given their many professional responsibilities (preparing course content, conducting research, academic service, etc.) [8]. This may be one reason students widely report dissatisfaction with the written feedback actually received from instructors [27].

To supplement their feedback, instructors teach students to seek feedback from potential end users who might use the design solution [12, 13]. For example, students might seek people with dietary restrictions to give feedback on the design of an app that filters restaurants based on the menu items. However, it can be difficult for students to quickly locate users who match the target audience, and who can commit the time needed to give feedback on the project [5].

Classroom Peer Feedback

Peer feedback asks people similar in ability to assess each other's in-progress work [23]. Peer feedback is scalable and also offers learning benefits. For example, it supports skill assessment and learning through exposure to different solution strategies [14, 18, 28, 31] and can lead to improved solution quality [9, 29]. Given its benefits, educators have applied peer feedback in a variety of classroom settings, from small studio courses to large online courses [23]. Researchers continue to improve the quality of peer review by studying interface features of the review platforms [15], rubrics [23], and characteristics of the review content [31].

However, peer feedback may be influenced by friendship, competition [36], and course workload [22]. Feedback could also be narrow due to students' exposure to the same learning material and their similar demographics [11, 17]. These limitations could affect the perceived quality of the feedback and students' willingness to act on it. It is therefore unclear how students would perceive feedback generated by classroom peers relative to external audiences.

External Crowd Feedback in the Classroom

Another alternative for scaling feedback generation is to have students solicit feedback from external crowds, such as social media (e.g. Facebook), online communities (e.g. Reddit), and paid task markets (e.g. MTurk) [17]. External crowds can quickly provide personalized feedback on many projects in parallel [5, 17]. Students may perceive the feedback favorably due to its authenticity [33] and potential insights outside the course material [17]. Being responsible for knowing when and how to gather feedback can also help students learn to be more self-directed in their work [7].

Researchers have begun to assess feedback from external crowds in the classroom. Xu et al. showed that students can apply the feedback from a task market to make non-trivial revisions to their graphic designs [40]. Other work has shown that the feedback from a task market can be of high quality and is perceived to improve the design process [25].

Hui et al. showed students can leverage social media to receive feedback and improve their work [16]. However, getting feedback from external crowds can require social capital [30], community membership, or payment [20]. Students may also be reluctant to share work online [16].

Our study contributes to this thread of work by providing empirical knowledge of how the *source* of the feedback (classroom peers or external crowds) affects its content, quantity, perceived quality, and actions taken.

External Crowd Feedback Outside the Classroom

Yen et al. [41] performed the only study we are aware of that compared the design feedback from different external crowds, including online communities, social media, and task markets. The study found that the feedback from different sources was of similar perceived quality; however, online communities offered more process feedback, task markets offered more design suggestions, and social media offered the most responses without payment. Most participants in the study had professional design experience.

Our study extends this prior work in two ways. First, Yen et al compared feedback between external crowds, whereas our study compares external crowd feedback to that of classroom peers. It is critical for instructors to know how the feedback differs between these sources so that they can choose the source best suited to their students' projects. Second, our study generates the external feedback using the same platforms as in [41] but in the context of a course where students are learning about design and have little or no prior design experience. Our focus on students differs from the prior study which had more experienced designers.

RESEARCH QUESTIONS

Our study was designed to answer the following questions:

RQ1: How does the source (peers and external crowds) and stage of the design (early vs. late) affect students' **(a)** perceived value of and **(b)** action taken on the feedback?

RQ2: How do the source of the feedback and stage of the design affect the content of the design feedback generated?

RQ3: What do students perceive as the benefits and limits of leveraging peers and external audiences for feedback?

These questions were posed to help design educators understand how various sources of feedback could be effectively leveraged at different stages of the process. For example, external crowds may be more detail-oriented, which may be less helpful for early-stage prototypes. Answers to these questions will also provide a deeper understanding of how crowds driven by different incentives provide different types of feedback, which instructors can take advantage of when choosing which feedback.

METHOD

To answer these research questions, we conducted a study in a project-based user interface (UI) design course at the University of Illinois, a large public university in the U.S.

Design Course and Projects

The UI design course targeted upper-level undergraduate and beginning graduate students in computer science, but about 10% of the students came from engineering, psychology, and the arts. For most students, this was their first course on UI design. The instructor organized students into teams (4-6 students) to balance skill sets. There were 188 students (44 female) and 40 teams in the course. The lecture topics included user research, ideation, prototyping, and evaluation. The teams applied these topics to a UI design project of their choice. Example projects included a mobile route-finding app, a web app to render online news in the style of a print newspaper, and an app for managing vacation packing lists.

The projects were structured as a design process. Each stage corresponded to a project deliverable submitted for a grade. Two key deliverables were targeted in this study: *early*- and *lat*e-*stage* prototypes. An early-stage prototype consisted of drawn sketches representing the direction of the proposed project solution. A late-stage prototype was a programming implementation of a team's revised early stage design.

These two design stages represented major milestones and would especially benefit from formative feedback. Early-stage prototypes may benefit from feedback about goals, scope and approach; late-stage prototypes may benefit from feedback about implementation choices and aesthetics; and both may benefit from perspectives beyond the course staff. At each stage, teams received feedback from classroom peers and the external crowds. The feedback was generated as part of a course assignment. All students were required to complete the assignment at each stage, since learning to gather feedback online and write constructive feedback are important learning goals [1, 10]. Other project deliverables such as the proposals and user research reports received feedback only from the course staff. We adapted an existing online review platform (described in [41]) to collect and present feedback from each source and design stage. Figure 1 shows an example of a project team's early- and late-stage prototypes and some of the feedback they received.

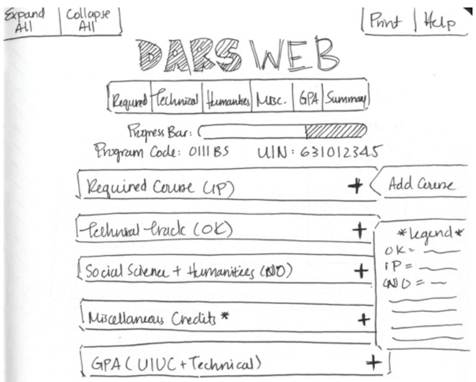
Peer and External Crowd Feedback

Project teams received feedback from four sources:

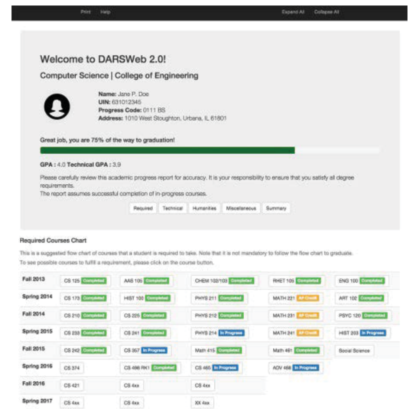
Peers: Each student was randomly assigned a design from another project team for feedback. Each team therefore received feedback from three to six peers at each stage. The links to the feedback forms were sent by email from the course staff as part of an individual assignment. Students entered their feedback on our platform and submitted it for grading by the course staff on a two-point scale: specific and helpful (2), too generic or short (1), or none given (0). Students were aware of the rubric prior to giving feedback.

Online community: Each project team posted the feedback link for their design to an online community of their choice, with instructions to target a relevant audience. This was performed as a group task to reduce a wave of requests to

Early-stage Prototype



Late-stage Prototype



Peers: "It looks really convenient! I like the idea of having the flexibility to add courses and check if they fulfil requirements. I also like the color coding feature. One thing that is not clear in the image is whether there will be a percentage indicator on the progress bar to get an accurate idea of the progress. Also, in my opinion, the buttons Expand All and Collapse All should be on the right side '+' button is located on the right side too. The Print and Help button can be on the left side. You could also add the option to save the report as a PDF."

Online Communities: "I like the idea you guys have implemented. A more UI friendly way of viewing your academic progress would definitely be a welcome development for us students. One thing I do want to say is that with your idea, obtaining information is going to be very tedious. I am sure you don't currently have a way to access each student's academic record. As a result, I assume, you would request all of this information from the user themselves…"

Task Market: "I think you have a great concept going. When you do finally add colors, I would add bright colors to make the page really pop."

Social Media: "Great work! Looks clean and interface looks user-friendly."

Peers: "Should the required courses be divided by year? I don't imagine courses need to be taken at particular times. Also I'm seeing the Required, Miscellaneous and Summary sections but where are the Technical and Humanities requirements? Is there a general overview of the courses taken already? seems like only a portion of the completed classes are displayed in the summary area. I do appreciate the clarity of the design, though it seems to be missing some information I think would really add to it."

Online Communities: "You might want to signify the number of hours a certain class fills and/or completed during each semester. Also keep in mind that some people complete classes over winter or summer semesters."

Task Market: "Overall, it's pretty clear. I would like to see all the boxes the same size. Maybe use different colors to differentiate courses too."

Social Media: "Your navbar is pushed in compared to your divs. This kinda looks asymmetrical / pyramid-like. I'd rather it be square."

Figure 1: An example of a project team's early- and late-stage prototypes for an online course planning interface and the anonymous feedback received from the four sources studied. Some of the feedback was omitted for brevity.

any one community. As a default, the course staff suggested the Reddit forum /r/design_critiques, a popular resource for online design critique. The team submitted a screen shot of the post and rationale for the community selected.

Social network: Each project team member posted the feedback link for their design to an online social network (Facebook or Twitter) using their own account. This was an individual task to smooth differences in each team's total network size [16]. Students submitted a screenshot as proof of their post. Students were not required to perform this task if they did not have a social media account or were uncomfortable using it for this purpose.

Paid task market: The research team posted the link to each team's design to Amazon.com's Mechanical Turk (MTurk) for feedback. Five pieces of feedback were solicited for each design to be consistent with the volume of peer responses. The feedback tasks paid $0.40, which, based on pilot data, equated to a wage at or above U.S. minimum wage. The task prompt stated: "*We are collecting feedback on a [design stage] design for a user interface. The user interface design project is part of a course project at a university.*" The research team posted all feedback tasks for the teams to eliminate the need for students to create accounts on MTurk and to have consistency in the configuration of the MTurk tasks across each project team.

All feedback was left anonymously, though the platform logged the source. We chose these sources because they are readily available and because they differ in terms of the incentives for giving feedback – course credit (peers), topical interest (online communities), social relations (social media), and financial gain (paid task market).

Procedure

The procedure consisted of two phases summarized in Figure 2. The first phase began two weeks before the due date for the early-stage prototypes, at the midpoint of the course. The corresponding homework assignment generated feedback from all four sources for early-stage prototypes.

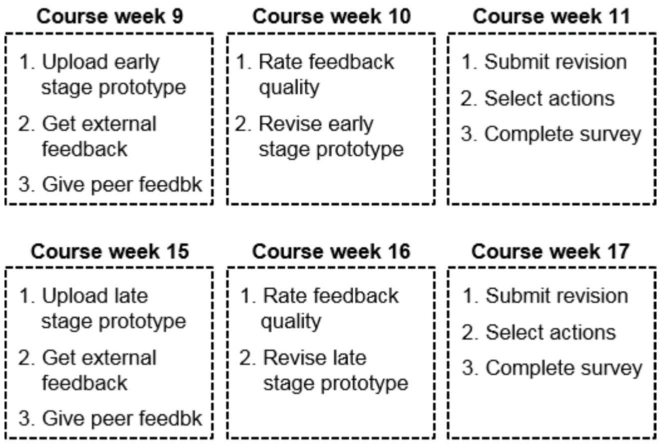


Figure 2: Summary of the nine-week experimental procedure. The early-stage prototype phase began in course week 9 (top row). The late-stage prototype phase began in course week 15 (bottom row). Prototypes were created in the weeks preceding the respective phase of the experiment.

For the assignment, each team uploaded an image or scan of their early-stage prototype to our platform along with a title and description. The platform created a link to the feedback form for the design and then stepped the team through posting the link to an online community and social network. Meanwhile, the research team posted the link to the paid task market for feedback. Students received individual emails with the links for giving peer feedback. Feedback generation lasted one week and all feedback was aggregated on our platform.

Teams had one week to iterate on the design after receiving feedback. After iterating, teams returned to our platform, selected the action taken for each feedback response, and completed a feedback reflection form (see Measures). The revised early-stage prototype was then submitted as part of the project requirements, completing the assignment.

For the second phase, the same procedure was repeated for the late stage prototypes. These prototypes were uploaded as a screenshot of the implemented user interface. This phase occurred near the final weeks of the course.

All students participated in the team projects and completed the two assignments as part of the course. At the end of the course, we distributed a consent form to the students. Only students who gave consent to use their data were included in our analysis. The study required coordination between the research team and course staff, which had no overlap. The study was approved by the IRB at our institution.

Measures

We measured the perceived quality, action taken, valence, content category, and length of each feedback response. We also collected demographics from the providers: gender, age, design expertise, and their reason for leaving feedback.

Teams collectively rated perceived quality of each feedback response on a 7-point scale (7=highest). After iterating on their design, teams selected the action taken on each piece of feedback: implemented fully, implemented partially, thought about it, or ignored it. All feedback was presented blind to condition to control for biases toward the feedback sources. These measures allowed us to gauge not only perceptions of the feedback, but also how it was applied.

The valence of feedback indicates its affective tone [34]. Given the size of the data set, we recruited two judges to rate the valence of each response on a 5-point Likert scale (1=Very Negative, 3=Neutral, 5=Very Positive). Judges were recruited from MTurk and were paid $0.12 per task. For a majority (56%) of the responses, judges agreed on the rating, which became the rating assigned. Of the disagreements, 82% of the responses differed by only one Likert scale point. In this and the remaining cases, a member of the research team cast the deciding vote.

To categorize content, the responses were partitioned into idea units, and the units were categorized using a taxonomy of critique discourse (e.g. judgement, brainstorming, and process) [4]. Details of the categorization are in Results. The word count of each response was also measured.

At the end of each design stage, teams completed a feedback reflection form. Teams rated the overall quality of the feedback received on a 7-point scale (1=Low, 7=High) and their comfort level sharing their design with each of the three external crowds (1=Very Uncomfortable, 7=Very Comfortable). The form also asked teams to select which source they expected to provide the highest quality feedback (or indicate that they expected the same quality from all sources) and to write a free response characterizing the strengths and weaknesses of the feedback.

Student Interviews

At the completion of the course, we conducted 45-minute semi-structured interviews with 10 students. Participants were recruited through a message distributed to the course roster. Each participant got $10 as remuneration. Interviews were audio recorded and then transcribed for analysis.

Interview questions focused on perceptions of soliciting the feedback from external crowds ("What do you think is the largest benefit [and weakness] of getting feedback from external crowds?") and which method they would prefer in a future course ("Which approach [peers or external crowds] would you prefer to leverage in similar engineering design courses?"). We used a bottom-up analysis to code emergent themes in the interview data [3]

RESULTS

Our study examined how the source (peers and external crowds) and design stage (early vs. late) affect students' perceived value of and action taken on the feedback received (RQ1); how the source and design stage affects the content of the feedback (RQ2); and student perceptions of receiving feedback from peers and external crowds (RQ3).

	OC	TM	SM	Peers
Early stage	34	199	280	93
Late stage	39	211	127	94
Total	73	410	407	187

Table 1: Number of feedback responses per source and design stage. OC = Online Communities, TM = Task Market, SM = Social Media, Peers = Classroom Peers.

Quantity of Feedback

We collected 1159 feedback responses across all sources and design stages. In total, 127 out of 188 peers in the course consented to participate in the study, a 68% response rate. We removed ten of the feedback responses due to inappropriate content. The remaining responses were filtered for double submissions (50) and omissions (32), leaving 1077 responses for analysis. The distribution of responses between source and design stage is summarized in Table 1. This pattern was consistent across both design stages. However, the number of social media responses was greater in the early stage (280) than in the late stage (127) (χ^2=113.5, p<0.001). This may reflect the social cost of requesting feedback from one's online social network [30].

Peer Feedback Has Higher Perceived Quality (RQ1a)

Figure 3 summarizes students' perceived quality ratings. We performed an ANCOVA with Source (peers vs. external crowds) and Stage (early vs. late) as factors. The provider's self-reported gender, age category, and design expertise (1-5, 5=expert) were added as co-variates. The ANCOVA showed that feedback from peers (μ=5.03) was rated higher than from the external crowds (μ=4.10; F(1,1065)=38.1; p<0.001). As shown in Figure 3, peer feedback was rated about 20% higher than that of external crowds. One explanation is that peers were more likely to identify issues relevant to the learning goals of the course. For instance, during an interview, one student stated, "*the peer source [addressed] issues we learned about in class*".

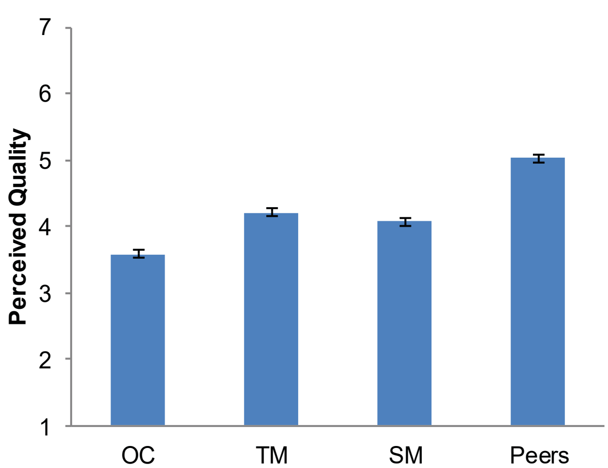


Figure 3: The means of the perceived quality ratings. OC = Online Community, TM = Task Market, SM = Social Media. Error bars show pooled standard error (0.06).

Stage did not have a main effect, there were no interactions, and no effects of interest were found for the covariates.

Pairwise comparisons using Tukey's method showed that feedback from peers (μ=5.03) was rated higher than from online communities (μ=3.6, p<0.001), social media (μ=4.08, p<0.001), and the task market (μ=4.20, p<0.001). Feedback from online communities was perceived to be the lowest quality, which may be due to the cursory nature of the feedback typically received from this source [38, 39]. No other differences were detected for perceived quality.

Peer Feedback is Implemented More (RQ1b)

We compared the counts of the project teams' selected actions on the feedback using Source and Stage as factors. Feedback responses without a selected action (n=19) were removed from the analysis. Peer feedback was implemented more (23% *implement partially*), thought about more (34% *think about it*) and ignored less (14% *ignore*) relative to the feedback from external crowds (16% *implement partially*, χ^2=5.2, p=0.022; 22% *think about it*, χ^2= 12.0, p<0.001; and 35% *ignore*, χ^2=28.8, p<0.001). There were no other differences detected, and no differences between stages.

Peers Write Longer Feedback

We analyzed the length of the feedback using the same model as for perceived quality, except with length as the dependent measure. The ANCOVA showed an interaction effect (F(1,1065)=19.04, p<0.001): feedback from peers and online communities was longer for early-stage designs (μ=108 and 74 words, respectively) than for late-stage designs (u=70 and 46, respectively; p<0.001 in both cases). Feedback from social media and the task market was similar in length between stages. Social media feedback had a mean length of 40 words at the early stage and 45 words at the late stage, while task market feedback had a mean of 43 words for the early stage and 47 words for the late stage. The ANCOVA also showed that peers wrote more content (μ=89 words) than providers in online communities (μ=59), the task market (μ=45), and social media (μ=43) (p<0.001 all cases). No other covariates of interest were significant. A regression analysis showed a small correlation between length and perceived quality (R^2=0.14, p<0.001), corroborating that students value longer feedback [42].

Valence is More Negative for Late-Stage Prototypes

The valence of the feedback was analyzed using the same model as for perceived quality, with valence as the dependent measure. The ANCOVA showed a main effect of Stage (F(1,1065)=4.34, p=0.037): feedback at the late stage (μ=2.92) exhibited more negative valence than at the early stage (μ=3.04, p=0.036). This may be because late-stage prototypes have specific graphic design elements that elicit specific critiques [37]. Source did not have a main effect, and none of the covariates had a significant effect.

Content Focuses on Assessment and Analysis (RQ2)

To categorize the content, we first partitioned the feedback in our dataset into idea units. An idea unit represents a coherent unit of thought and may be comprised of a phrase,

Category	Definition	Example Idea Units
Judgement	Critic reacts to what they see and renders some assessment of its quality.	"I like the color scheme."
Process-Oriented	Critic makes statements or asks questions about the process that students might have used or could use to create the design.	"Use animations for more polished prototypes."
Brainstorming	Critic asks questions or makes rhetorical statements about future imagined possibilities for the design.	"Also, this would be great for use within Airports for travelers."
Interpretation	Critic reacts to what they see and tries to make sense of the concept or product.	"I have no idea what is going on with this page, it doesn't make any sense."
Direct Recommendation	Critic gives specific advice about a particular aspect of design. Feedback is focused, purposeful, and specific.	"The picture could use annotations on the controls for screen cropping."
Investigation	Critic requests information (typically by questioning) about the design or the design process.	"What other information will you include to help students take their desired courses?"
Free Association	Critic makes reactive, associative statements about the design.	"it doesn't seem that different than similar products right now."
Comparison	Critic contrasts the design or design process with something else in a focused, intentional way.	"Though I can see this is a preliminary design, it is very similar to DropBox -- how do you plan on differentiating it?"
Identity-Invoking	Critic makes statements or asks questions to suggest that students consider the larger picture of themselves as designers in a future professional community.	"seems you have learned little or nothing from [the design course]."
Support	Generic praise and support of the designer, not the design.	"keep up the great work guys!!!."

Table 2: Categories used for feedback content analysis, adapted from [4].

sentence, or group of sentences. The partitioning yielded a total of 3116 idea units; 282 from online communities, 1046 from MTurk, 1022 from social media, and 766 from peers.

Given the size of our dataset and the unequal distribution of the responses, we sampled the idea units to ensure that each source was equally represented. Online communities gave the fewest responses, with 282 idea units. Therefore, we randomly sampled 282 idea units from each of the other sources, giving a total of 1128 (282 x 4) units for analysis.

Idea units were categorized using a taxonomy of critique discourse in design [4]. The taxonomy has nine categories: *judgement, process-oriented, brainstorming, interpretation, direct recommendation, investigation, free association, comparison*, and *identity-invoking*. After reviewing the data set, we added a category for *support* (the idea unit offered praise or encouragement). Definitions and examples of each category are shown in Table 2.

We trained two independent labelers outside our research team on 10% of the dataset (n=106 idea units). After several iterations, Krippendorff's α, a measure of inter-rater reliability [21], did not exceed 0.6. Thus, both labelers categorized all idea units independently. We aggregated the resulting category assignments, and the labelers worked together to resolve disagreement. If agreement was not reached, a member of our research team cast the final vote.

The distribution of categories across the sources and stages is summarized in Table 3. The most noticeable pattern was that the vast majority (90%) of idea units fell into four categories: *direct recommendation (40%)*, *judgement (31%), interpretation (12%),* and *investigation (7%)*. This was consistent across sources and design stages. These four categories are most useful for the late stage, when there is more emphasis on the design representation and its quality.

However, these categories may be less suitable for the early stage. At this stage, students may benefit most from reflecting on how they approached the design, how the design compares to other solutions, and how it fits in a broader social and cultural context [37]. Feedback in the categories *brainstorming (3%), comparison (3%)*, *process-oriented (2%), free association (2%),* and *identity-invoking (<1%)* would best speak to those issues, but these categories were rarely addressed by any of the four sources.

Student Perceptions (RQ3)

In this section, we report results from the project team surveys collected at the early stage (n=40) and late stage (n=40) and the individual interviews (n=10). Quotations are reported with a T for team survey responses and an S for individual student interview responses.

Peers Expected to Be Highest Quality Source

Teams selected which source they expected to provide the highest quality feedback. Expectations for the highest quality source differed at both the early stage (χ^2=16.0, p=0.0031) and late stage (χ^2=29.7, p<0.001). Overall, 45% of teams expected peer feedback to be of highest quality and 40% expected external crowds to be of highest quality. The remaining 15% expected them to be the same quality.

At the late stage, more teams expected peer feedback to be of highest quality (50%) than online communities (18%, p= 0.0460), the task market (8%, p=0.005), and social media (15%, p=0.019). At the early stage, teams' preferences were not as strong, with the only difference being between peer feedback (40%) and the task market (5%, p=0.005).

	Early-Stage Prototypes				Late-Stage Prototypes			
	OC	TM	SM	Peers	OC	TM	SM	Peers
Judgement	38 (29%)	38 (31%)	53 (29%)	48 (29%)	53 (36%)	52 (32%)	26 (29%)	32 (29%)
Process-Oriented	3 (2%)	2 (2%)	4 (2%)	2 (1%)	1 (1%)	4 (2%)	0 (0%)	2 (2%)
Brainstorming	2 (2%)	5 (4%)	6 (3%)	2 (1%)	5 (3%)	6 (4%)	1 (1%)	1 (1%)
Interpretation	21 (16%)	8 (6%)	24 (13%)	29 (17%)	13 (9%)	14 (9%)	10 (11%)	19 (17%)
Direct Recommendation	43 (33%)	61 (49%)	73 (39%)	59 (35%)	48 (33%)	76 (47%)	42 (46%)	40 (36%)
Investigation	16 (12%)	2 (2%)	10 (5%)	18 (11%)	8 (5%)	2 (1%)	5 (5%)	16 (14%)
Free Association	2 (2%)	4 (3%)	0 (0%)	2 (1%)	6 (4%)	3 (2%)	1 (1%)	0 (0%)
Comparison	4 (3%)	2 (2%)	7 (4%)	8 (5%)	6 (4%)	3 (2%)	2 (2%)	1 (1%)
Identity-Invoking	1 (1%)	0 (0%)	0 (0%)	0 (0%)	2 (1%)	0 (0%)	1 (1%)	0 (0%)
Support	0 (0%)	1 (1%)	6 (3%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)

Table 3: The distribution of occurrences across critique categories for each source and design stage. OC = Online Communities, TM = Task Market, and SM = Social Media. Percents do not sum to 100 because some idea units did not fit into any category.

Students Somewhat Comfortable Sharing Designs Online

Project teams rated their comfort level sharing their designs with each external crowd (1=Very Uncomfortable, 7=Very Comfortable). Students reported being neutral to somewhat comfortable sharing their designs online. The distributions were μ=5.0, σ=1.90 for online communities; μ=4.2, σ=1.99 for social media; and μ=4.0, σ=1.60 for the task market.

A repeated measures ANOVA with Source and Stage as factors and level of comfort rating as the measure showed that Source had a main effect ($F(2,197)=7.55$, $p=0.0007$). Post-hoc tests revealed that teams were less comfortable with having their designs shared on the task market (μ=4.0) than with online communities (μ=5.0, $p=0.0047$). This difference may be due to students being less familiar with the task market, since the researchers posted the designs on their behalf. There was no difference detected in comfort level between the early- and late-stage prototypes.

Strengths and Weaknesses of the Feedback

Project teams were asked to characterize the strengths and weaknesses of the feedback they received at both stages. Since students were blind to condition, they reported on the feedback holistically. We discuss the responses in aggregate since they were similar between the two stages. We use n to indicate the number of responses citing a given theme.

The most commonly cited strengths of the feedback were that it brought new issues to the team's attention through specific, useful suggestions (n=28), it was more honest and unfiltered due to anonymity (n=6), and it produced diverse perspectives by including access to broader audiences (n=5). A team building a web-based weather application described the usefulness of specific suggestions:

> *"Having people outside our groups reviewing our design, we are able to receive some insights into things we previously overlooked. For example, we originally did not think that precipitation was important, but now we know that most people look at precipitation when they want to find out the weather."* (T21)

Students valued not only specific feedback, but also feedback based on diverse perspectives. A team working on a mobile route finding app explained:

> *"The strength is that we learned a lot from people with different background and opinions from different perspectives. A lot of them are quite useful."* (T11)

Students also valued the honesty of the feedback. A team working on a multi-thread messaging platform noted:

> "*Since the feedback is anonymous, people can very easily troll you. …However, the same anonymity allows people to provide honest feedback, which is totally worth the tradeoff."* (T2)

While students appreciated many aspects of the feedback, they struggled with vague or irrelevant content (n=28). For instance, teams building a mobile restaurant finding app and a course management system for university students said, respectively:

> *Very vague feedback. Some of it was not even feedback.*" (T6)

> *"A lot of feedback was just 'good job' with no actual suggestions for improving the platform."* (T15)

This also manifested for some teams as a low signal to noise ratio. One team working on a mobile home meal sharing app explained:

> "*While it was good to hear the compliments about the app, it was also a lot to filter through to find the main design points that were important to change."* (T19)

Teams also reported frustration when the providers appeared to misunderstand the context of the design (*n*=10). For example, one team developing a mobile work scheduling app stated:

> "*Generally a lot of the feedback was questions about how the prototype worked as opposed to giving substantial feedback.*" (T37)

Students See Tradeoffs for Different Feedback Sources

Interviewees (9 of 10) believed the choice between all of the sources of design feedback depended on the source's knowledge of the topic and design domain, and how well the source represented the target user audience. Interviewees perceived the non-expert nature of external crowds as both a strength and a weakness. When non-expertise was viewed as a strength, interviewees pointed to situations where external crowds could provide fresh insights. For example, one interviewee remarked that having external crowds critique his team's re-design of an online course management tool was useful due to their novel perspective:

> "*Since we had a student focused design, to have people that are possibly not students, like from Reddit, was one of the greatest strengths of it…because they point out a lot of things that we didn't notice, because with a student, they know how Compass runs, so they're looking for the same functions in our design. But someone who hasn't used it, it's completely new to them.*" (S2). Another interviewee shared a similar view: *"Sometimes [external crowds] will provide some more feedback…from a non-designer's perspective, from practical use.*" (S5)

In other situations, non-expertise was seen as a weakness. One interviewee said,

> "*The purpose of the design is to make a design that's going to function the best. I want someone who could help me with functionality as opposed to just looks.*" (S2)

In these cases, peers or online communities were seen as the most expert source. Another interviewee explained why:

> *"For the peers, they will give more technical advice… The people who are on Reddit, they might be more expert, and they focus more on the details than our peers-our friends, for example. So, they might give more advanced ideas or constructive ideas."* (S6)

Finally, interviewees felt it would be most helpful to get feedback from the source most representative of their target audience. For instance, one interviewee's group was designing an interface for chemistry lab equipment and felt they needed feedback from this specialized user audience:

> *"It depends on what the functionality of the project is and the target users. For myself, I have a project where we're designing a switch...for a chemistry lab…the users are mostly scientists, [so] the crowd user may not be really helpful."* (S1)

DISCUSSION AND FUTURE WORK

A goal of our study was to compare the feedback from classroom peers and external crowds at different design stages for student projects. Our results showed that peers produced feedback that was of higher perceived quality, was acted on more, and was longer than the feedback from the external crowds. Students also expected the feedback from peers to be of higher quality than from any of the external crowds. One reason for this pattern was that even though peers were anonymous to each other, they were accountable to the course staff for completing the feedback assignments. Though the feedback assignments were only part of the coursework, our experience as instructors is that receiving a grade is a strong incentive for students to perform the work. Students also had incentives that overlapped with the external crowds: classroom peers typically have some social ties with each other, and given that this was an elective course, many students registered due to an interest in the topic.

Our results showed that crowd feedback is still a viable option for student projects since the feedback was perceived to be of comparable quality (μ=4.1 vs. μ=5.0 for peers). External crowd feedback could therefore supplement peer feedback received at key project milestones or replace it when tight schedules or an already high student workload make peer feedback less desirable. Instructors could also use external crowds to help students learn to become more self-directed in planning for and acquiring feedback [7], as this is an essential skill for innovation careers [11]. Finally, external crowds can be helpful for accessing specialized user audiences, as mentioned by student interviewees.

All sources produced feedback that primarily spanned four of the nine categories of the critique taxonomy: suggestions for improvement (*direct recommendation*), assessments of quality (*judgement*), and sense-making statements (*investigation* and *interpretation*). These categories are especially beneficial for designs at the late stage [37]. The other categories (e.g., *brainstorm*, *process*, and *comparison*) may be more suitable for early-stage designs because the content would likely address how teams approached the design problem, how the solution path compares to alternatives, and how it fits in a larger context. To fill these categories, instructors may need to scaffold the feedback generation process with appropriate prompts (e.g. as in [26]) or provide this type of feedback themselves.

The design stage had minimal effects in our study. With Stage as a factor, our analysis did not detect differences in perceived quality of the feedback, actions taken, content categories, or comfort sharing, and showed only minimal effects for expectation of quality. Instructors may therefore weigh the stage of the design less than other factors when considering which sources to use for feedback.

Students reported feeling comfortable sharing their designs (all means > 4 on a 7-point scale) with all three external crowds. This may be due the anonymous feedback context our experimental design created. Due to an omission in our survey, we did not collect comfortableness ratings for sharing designs with classroom peers.

In our study, students benefited from the work performed by feedback providers recruited from online communities and social media, but did not reciprocate that effort. In the future, if an instructor wishes to leverage these platforms, s/he should consider requiring students to pay forward that effort. This may also help students understand how to approach these platforms when later requesting feedback.

Our results for external crowds can be compared to a study by Yen et al that used similar platforms to generate feedback for design projects outside of a classroom context [41]. Both studies found that average perceived quality for all crowds was just above 4 on a 7-point scale with minimal or no differences between crowds. For the crowds where the number of responses were not controllable, Yen et al observed that online communities produced more responses than social networks, whereas our results showed the opposite pattern. In our study, student teams utilized their collective social networks, as in Hui et al. [16], which may have increased responses due to a larger network size. Also, all the feedback requests were posted to online communities in parallel, which may have caused feedback providers to give less attention to each posted design. Using the same taxonomy, both studies found that *judgements* and *direct recommendations* were most common in the feedback.

Our results also produced two main design implications for improving the effectiveness of crowd-based platforms for feedback exchange. First, the prototypes in our study were represented on our platform as a single image. This approach supports any implementation (e.g., mobile, desktop, and Web), and mimics existing feedback platforms (e.g., Reddit). However, students reported that this constraint caused the feedback to focus more on aesthetics rather than usability at the late stage. Enabling richer design representations, such as design images organized into tasks or videos of prototype use, could help focus feedback on other important aspects of the late stage design. Second, students mentioned the desire to have a dialogue with their feedback providers. Enabling this dialogue during feedback generation could help clarify the issues raised, answer questions, and promote a longer-term relationship between the providers and designers throughout the project.

In addition to the issues already discussed, we see several opportunities for future work. First, this paper viewed external crowds and peers as separate sources. Future work could examine how external crowds and peers could work together to produce feedback that is of higher quality than either could produce alone. Second, the external crowds in our study provided a large, broad audience from which to generate feedback. Future work could explore how online technologies might connect students with a smaller group of target users from external audiences for the duration of their design projects. Finally, as prior work has shown that feedback helps students learn [23], a future study should investigate how the feedback received from peers and external crowds affects learning and project outcomes.

LIMITATIONS

One limitation of our study is that we conducted it in a single university course. Future work is needed to test the generalizability of our findings in additional course settings. Second, we did not scaffold feedback generation, instead using a similar open-ended prompt for each source to be consistent across conditions. However, previous work has shown that scaffolding using rubrics [43], pre-authored statements [24], task decomposition [40], and framing features of the review platform [14] can improve feedback from non-experts. Future work should test how different scaffolding and framing techniques affect the pattern of results reported in this paper. Third, while limiting the use of peer feedback reduces student workload, we did not measure the additional student effort required to gather external crowd feedback. Future work comparing the relative effort required for each activity is needed to understand the extent to which crowd feedback reduces student workload. Finally, we measured feedback quality based on student perceptions. Future work should include additional measures such as expert evaluation to create a more complete profile of feedback quality.

CONCLUSION

This paper compares the feedback generated by classroom peers and three external crowds at different stages of the design process in a project-based design course. Our results show that while peers outperformed the external crowds on certain measures, such as perceived quality of the feedback and actions taken on it, getting feedback from external crowds is still a viable option when peer feedback is not desirable, and may be preferred for projects that target specialized audiences. Our analysis also points to the need to broaden the scope of the feedback (e.g. to include more process-oriented and conceptual feedback, and comparisons to existing design solutions) offered by all four sources. Our results will help instructors know how to better coordinate the use of classroom peers and external crowds in a way that is most beneficial to learning and student projects at different stages of the design process.

ACKNOWLEDGEMENTS

This work was funded in part by the National Science Foundation under awards 1530818, 1462693, 1122206 and 1122320, the first author’s NSF Graduate Research Fellowship under award 1144245, and an award received by the last author from the Grants for the Advancement of Teaching Engineering program at the University of Illinois.

REFERENCES

1. Phyllis C. Blumenfeld, Elliot Soloway, Ronald W. Marx, Joseph S. Krajcik, Mark Guzdial and Annemarie Palincsar. 1991. Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26

(3-4). 369-398. http://dx.doi.org/10.1080/00461520.1991.9653139

2. Donald J. Boyd. 2015. Public Research Universities: Changes in State Funding. *American Academy of Arts and Sciences*.
3. Juliet Corbin and Anselm Strauss. 2014. *Basics of Qualitative Research; Techniques and Procedures for Developing Grounded Theory*. Sage Publications.
4. Deanna P. Dannels and Kelly N. Martin. 2008. Critiquing Critiques: A Genre Analysis of Feedback Across Novice to Expert Design Studios. *Journal of Business and Technical Communication*, 22 (2). 135-159. http://dx.doi.org/10.1177/1050651907311923
5. Steven P. Dow, Elizabeth Gerber and Audris Wong. 2013. A Pilot Study of Using Crowds in the Classroom. in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 227-236. http://dx.doi.org/10.1145/2470654.2470686
6. Matthew Easterday, Daniel Rees-Lewis, Colin Fitzpatrick and Elizabeth M. Gerber. 2014. Computer supported novice group critique. in *Proceedings of the ACM Conference on Designing Interactive Systems*, ACM, 405-414. http://dx.doi.org/10.1145/2598510.2600889
7. Matthew Easterday, Daniel Rees-Lewis and Elizabeth Gerber. 2013. Formative feedback in Digital Lofts: Learning environments for real world innovation. in *AIED Workshops*.
8. Peggy Ertmer, Jennifer C. Richardson, Brian Belland, Denise Camin, Patrick Connolly, Glen Coulthard, Kimfong Lei and Christopher Mong. 2007. Using peer feedback to enhance the quality of student online postings: An exploratory study. *Journal of Computer-Mediated Communication*, 12 (2). 412-433. http://dx.doi.org/10.1111/j.1083-6101.2007.00331.x
9. Edmund Burke Feldman. 1981. *Varieties of Visual Experience: Art as Image and Idea*. Prentice-Hall, Englewood Cliffs, N.J.
10. Sebastian K. Fixson. 2009. Teaching Innovation through Interdisciplinary Courses and Programmes in Product Design and Development: An Analysis at 16 US Schools. *Creativity & Innovation Management*, 18 (3). 199-208. http://dx.doi.org/10.1111/j.1467-8691.2009.00523.x
11. Elizabeth M. Gerber, Jeanne Marie Olson and Rebecca L. D. Komarek. 2012. Extracurricular design-based learning: Preparing students for careers in innovation. *International Journal of Engineering Education*, 28 (2). 317-324.
12. John D. Gould. 1995. How to Design Usable Systems. in Baecker, R.M., Buxton, W., Grudin, J. and Greenberg, S. eds. *Readings in Human-Computer Interaction: Toward the Year 2000*, 93-121.
13. John D. Gould and C. Lewis. 1985. Designing for Usability: Key Principles and What Designers Think. *Communications of the ACM*, 28 (3). 300-311. http://dx.doi.org/10.1145/3166.3170
14. John Hattie and Helen Timperley. 2007. The power of feedback. *Review of Educational Research*, 77 (1). 81-112. http://dx.doi.org/10.3102/003465430298487
15. 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 *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 458-469. http://dx.doi.org/10.1145/2858036.2858195
16. 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*.
17. Julie S. Hui, Elizabeth M. Gerber and Steven P. Dow. 2014. Crowd-based design activities: helping students connect with users online. in *Proceedings of the ACM Conference on Designing Interactive Systems*, 875-884. http://dx.doi.org/10.1145/2598510.2598538
18. Christopher Hundhausen, Anukrati Agrawal, Dana Fairbrother and Michael Trevisan. 2009. Integrating pedagogical code reviews into a CS 1 course: an empirical study. in *Proceedings of the ACM Technical Symposium on Computer Science Education*, ACM, 291-295. 10.1145/1508865.1508972
19. Christopher D. Hundhausen, Dana Fairbrother and Marian Petre. 2012. An Empirical Study of the "Prototype Walkthrough": A Studio-Based Activity for HCI Education. *ACM Transactions on Computer-Human Interaction*, 19 (4). 1-36. http://dx.doi.org/10.1145/2395131.2395133
20. Aniket Kittur, Ed H. Chi and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. in *Proceedings of the ACM conference on Human Factors in Computing Systems*, ACM, 453-456. http://dx.doi.org/10.1145/1357054.1357127
21. Klaus Krippendorff. 2007. Computing Krippendorff's alpha reliability *Departmental papers (ASC)*.
22. Chinmay E Kulkarni, Michael S Bernstein and Scott R Klemmer. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. in *Proceedings of the Second ACM Conference on Learning@ Scale*, ACM, 75-84. http://dx.doi.org/10.1145/2724660.2724670
23. 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 Transactions on Computer-Human Interaction (TOCHI)*, 20 (6). 1-31. http://dx.doi.org/10.1145/2505057
24. Micah Lande and Larry Leifer. 2010. Difficulties student engineers face designing the future. *International Journal of Engineering Education*, 26 (2). 271.
25. 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 *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 473-485. http://dx.doi.org/10.1145/2675133.2675283
26. Xiaojuan Ma, Yu Li, Jodi Forlizzi and Steven P. Dow. 2015. Exiting the Design Studio: Leveraging Online Participants

for Early-Stage Design Feedback. in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 676-685. http://dx.doi.org/10.1145/2675133.2675174

27. David Nicol. 2010. From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35 (5). 501-517. http://dx.doi.org/10.1080/02602931003786559

28. Ken Reily, Pam Ludford Finnerty and Loren Terveen. 2009. Two Peers are Better Than One: Aggregating Peer Reviews for Computing Assignments is Surprisingly Accurate. in *Proceedings of the ACM International Conference on Supporting Group Work*, 115-124. http://dx.doi.org/10.1145/1531674.1531692

29. Howard Risatti. 1987. Art criticism in discipline-based art education. *Journal of Aesthetic Education*, 21 (2). 217-225. http://dx.doi.org/10.2307/3332751

30. Jeffrey M Rzeszotarski and Meredith Ringel Morris. 2014. Estimating the social costs of friendsourcing. in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, ACM, 2735-2744. http://dx.doi.org/10.1145/2556288.2557181

31. D Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional Science*, 18 (2). 119-144. http://dx.doi.org/10.1007/BF00117714

32. D. Royce Sadler. 2010. Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35 (5). 535-550. http://dx.doi.org/10.1080/02602930903541015

33. David Williamson Shaffer and Mitchel Resnick. 1999. "Thick" Authenticity: New Media and Authentic Learning. *Journal of Interactive Learning Research*, 10 (2). 195-215.

34. Mostafa Al Masum Shaikh, Helmut Prendinger and Mitsuru Ishizuka. 2008. Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. *Applied Artificial Intelligence*, 22 (6). 558-601. http://dx.doi.org/10.1080/08839510802226801

35. Valerie Shute. 2007. Focus on Formative Feedback. *Educational Testing Service*, Research report RR-07-11.

36. Holly Smith, Ali Cooper and Les Lancaster. 2002. Improving the Quality of Undergraduate Peer Assessment: A Case for Student and Staff Development. *Innovations in Education and Teaching International*, 39 (1). 71-81. http://dx.doi.org/10.1080/13558000110102904

37. Maryam Tohidi, William Buxton, Ronald Baecker and Abigail Sellen. 2006. Getting the right design and the design right. in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1243-1252. http://dx.doi.org/10.1145/1124772.1124960

38. Wesley Willett, Jeffrey Heer and Maneesh Agrawala. 2012. Strategies for crowdsourcing social data analysis. in *Proceedings of the ACM conference on Human Factors in Computing Systems*, ACM, 227-236. http://dx.doi.org/10.1145/2207676.2207709

39. Anbang Xu and Brian P. Bailey. 2012. What Do You Think? A Case Study of Benefit, Expectation, and Interaction in a Large Online Critique Community in *CSCW*, 295-304. http://dx.doi.org/10.1145/2145204.2145252

40. 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 *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 1637-1648. http://dx.doi.org/10.1145/2675133.2675140

41. 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 *Proceedings of the ACM Conference on Designing Interactive Systems*. http://dx.doi.org/10.1145/2901790.2901820

42. Alvin Yuan, Kurt Luther, Markus Krause, Sophie Vennix, Steven P Dow and Björn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. in *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1005-1017. http://dx.doi.org/10.1145/2818048.2819953

43. Haoqi Zhang, Matthew W. Easterday, Elizabeth M. Gerber, Daniel Rees Lewis and Leesha Maliakal. 2017. Agile Research Studios: Orchestrating Communities of Practice to Advance Research Training. in *Proceedings of Computer Supported Cooperative Work*, to appear. http://dx.doi.org/10.1145/2998181.2998199