Listen to Others, Listen to Yourself: Combining Feedback Review and Reflection to Improve Iterative Design

Yu-Chun Grace Yen
Computer Science
University of Illinois
Urbana, IL
yyen4@illinois.edu

Steven P. Dow
Cognitive Science
UC San Diego
San Diego, CA
spdow@ucsd.edu

Elizabeth Gerber
Segal Design Institute
Northwestern University
Evanston, IL
egerber@northwestern.edu

Brian P. Bailey
Computer Science
University of Illinois
Urbana, IL
bpbailey@illinois.edu

ABSTRACT
Feedback from diverse audiences can contain ambiguity and contradictions, making it difficult to interpret and act on. To promote deeper interpretation of feedback, we tested the effects of combining a reflection activity and reviewing external feedback for an iterative design task. Designers (N=90) created a design and revised it after a) performing a reflection activity *before* reviewing feedback, b) performing the reflection *after* reviewing feedback, c) performing the reflection only, or d) reviewing the feedback only. We measured design quality, depth of revision, perceived effort, and confidence; and categorized the content produced from the reflections. We found that performing reflection *after* feedback review led to the largest increase in perceived quality for the revised designs, and performing reflection and feedback review regardless of the order resulted in the most extensive revision. Our results also showed that performing the reflection alone yielded outcomes that were similar to when only reviewing feedback, and either activity led to better outcomes than the control condition (no feedback or reflection). Designers stated that the reflection helped them recall their goals, question their choices, and prioritize revisions. We argue that designers should perform a lightweight, explicit reflection to enhance their iterative process, and discuss implications for feedback platforms.

Author Keywords
Creativity; design; feedback; reflection; crowdsourcing.

ACM Classification Keywords
H.5.3 [Information interfaces and presentation]: Group and Organization Interfaces – Collaborative computing.

INTRODUCTION
Feedback is critical for performing creative design. It helps designers iterate on their in-progress solutions by revealing problems, gaining insight, and unblocking fixation [24]. However, the feedback received from diverse audiences such as those reached via social media and crowdsourcing [20, 32, 50, 52] can contain contradictions and ambiguity due to the providers having different motives [52], expertise [50] and perspectives [52]. Designers can therefore struggle to interpret and learn from the feedback, and act on it to improve their design solutions [7, 40, 50]. See Figure 1.

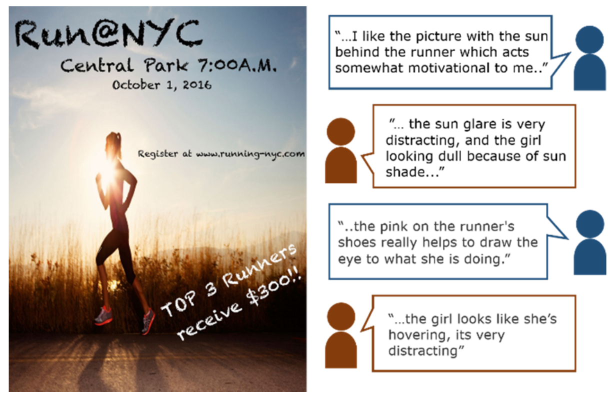

Figure 1. Example of a design and corresponding feedback generated from our study. The feedback exhibits conflict, thus requires deeper interpretation from the designers.

To help designers interpret feedback more effectively, we develop a lightweight reflection activity and test its placement relative to feedback review for iterative design. Reflection is a meta-cognitive process in which a designer assesses the design situation and its relation to the project goals [39]. Prior work has shown that performing reflection after feedback review improves management decisions [1] and learning outcomes [27]. The improvements have been attributed to the feedback review facilitating a more effective reflection [1]. Our research extends this prior work by testing the effects of placing the reflection either *after* or *before* feedback review. In the latter case, we speculate that the reflection could facilitate the effectiveness of the feedback review. Our research also extends the prior work to the context of creative design.

We conducted an online study in which designers (N=90) each created an initial graphic design for a public running event and then revised it after performing an activity in one of four conditions: *reflect-before-feedback*, *reflect-after-feedback, feedback-only*, and *reflect-only*. The study also included a control condition in which designers revised their designs without reviewing feedback or reflecting. The

C&C '17, June 27-30, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4403-6/17/06…$15.00
http://dx.doi.org/10.1145/3059454.3059468

participants in the study were mostly novice designers. We measured the perceived quality of the design and the depth of revision using ratings from domain experts and the designers themselves. Designers also rated their confidence and effort invested in the designs, and rated the usefulness of the activities performed. We also categorized the content produced from the reflection activity.

We found that reflection *after* feedback review led to the largest increase in the revised design quality according to the experts, and without designers over-estimating the amount of improvement they made. Designers reported a heightened focus on improving aesthetics such as layout, color, and fonts in this condition. If the reflection was performed *before* feedback review, we found that designers focused more on the concept of the design, but the quality of the revision did not increase as much as when the reflection was performed after feedback review. Performing the reflection and feedback review regardless of the order resulted in the most extensive revision. Our results also showed that performing the reflection activity alone yielded outcomes that were similar to when only reviewing external feedback, and either activity led to better performance compared to the control condition (no feedback or reflection). Designers reported that the reflection activity helped them recall their design goals, question their choices, and plan and prioritize their revisions.

Our work makes two contributions to the creativity and cognition community. First, we offer deeper empirical understanding of how ordering reflection and feedback review affects designers' performance and perceptions for a creative design task. Our results argue that designers should perform a simple, lightweight reflection activity to help ensure each design iteration is as productive as possible. Second, we offer implications for implementing a reflection activity in feedback platforms to encourage its adoption.

RELATED WORK

We discuss how our work builds on the theory of reflection in design. We also situate our contribution in the context of prior work that has combined reflection and feedback and other approaches for promoting feedback engagement.

Reflection in Creative Design

Schön distinguishes two categories of reflection: reflection-*on*-action, which refers to when a designer revisits prior design episodes, and engages in a meta-cognitive process to extract meaning from those experiences; and reflection-*in*-action, which refers to gaining insight while acting on a creative artifact [41]. The reflection activity developed in this paper is best characterized as reflection-*on*-action since it is an intentional act of writing an assessment of the design situation.

Schön's work has inspired many tools and methods to support reflection in design [3]. To support reflection-in-action, researchers have pursued approaches to increase the "talk-back" of design representations. Fischer and colleagues championed the approach of embedding software critics into a design tool [14]. The critics could analyze the design representation and generate feedback on-demand by applying domain-centric design principles [13]. Another approach has been to explore user interfaces for better representing design alternatives and the meaning of the design situation [22, 35]. A third approach has been to enable designers to generate design alternatives by having access to their design histories [9, 28, 29].

For reflection-on-action, researchers have explored the creation of visualizations of design processes [28, 34, 42]. The visualizations allow designers to notice patterns in their workflows, revisit memorable ideas, examples and decisions, and extract lessons that will shape how they approach future projects. Others have viewed reflection as an ongoing activity throughout the design process that can be regularly performed to sensitize designers and end users to each other's goals, knowledge, and motivations [45].

Our work differs from this corpus of prior work because we are testing how a reflection activity can be best coordinated with the review of feedback. Additionally, though reflection is widely recognized as a core practice in creative design [3], there is a paucity of literature that quantifies the effects of reflection on design. Our experiment adds to this literature by including a condition that isolates the effects of performing a reflection activity on design performance.

Using Reflection to Encourage Feedback Engagement

Feedback is critical for performing creative work because it helps a person assess the quality of her in-progress solution, discover its shortcomings, and gain insight for improving it [24]. However, the effectiveness of feedback on the design depends in large part on how well a person is able to interpret the content, learn from it, and formulate an effective action plan [5, 25, 33, 37]. Though feedback interpretation is a long-standing problem in many creative contexts, the issue is becoming more critical as people increasingly leverage online sources for gathering feedback on their creative work [21]. Feedback from online sources can be gathered fast, is affordable, and can tap a global audience [50]. Despite these benefits, feedback received online can be ambiguous, contradictory, and of variable quality because the providers typically have different motivations [52], expertise [50], and perspectives [52] for writing helpful feedback. These issues can cause the feedback to be especially difficult to interpret.

One approach to promote deeper engagement with feedback is to incorporate a reflection activity. For example, in [1], users responded to simulated work emails that involved different categories of tasks such as coordination, decision making and resource allocation. Users received feedback on their task performance, and then developed and wrote an action plan while users in a control condition did not. The results showed that performing the reflection activity after reviewing feedback improved task performance more than receiving feedback alone [1]. The improvement was likely due to the reflection prompting deeper processing of the feedback and task context, which allowed the lessons to be more effectively applied to a second task. Similarly, in an educational setting, students were required to read feedback on their solutions and submit an action plan before the scores were released. The results showed that this approach helped the students view feedback as a means for learning rather than solely a justification for their grades [27].

Our work shares the goal of using reflection to promote deeper processing of feedback. We extend prior work by testing how placing a reflection activity, either before or after reviewing feedback, affects an iterative design task. We also extend prior work by measuring perceptions of performance, in addition to the task outcomes, and by analyzing the content from the reflection activity.

Other Approaches for Feedback Engagement

Drawing from the broader literature on sense-making, there are approaches beyond reflection that could be applied to promote engagement with feedback. These approaches include applying tools for interactively visualizing and organizing the data [2, 26], scaffolding dialogue around the data [6], leveraging others' interpretations of the data [15, 36, 47], and increasing the transparency of the data creation process [30, 49]. We chose to test the integration of a reflection activity into the iterative design process because reflection has been long recognized as a core practice in creative design [3, 17] and learning [38, 43]. It can also be embedded in online communities and feedback platforms, or performed on one's own without additional tools.

RESEARCH QUESTIONS

The goal of our study was to test the effects of integrating an explicit reflection activity into the iterative design process. We focused on the following research questions:

RQ1: How does integrating a reflection activity into an iterative design process affect perceived design quality, degree of revision, and perceptions of design performance?

RQ2: How does the sequence in which the reflection activity is performed – either before or after reviewing external feedback – affect these same measures?

RQ3: What are the perceived benefits and limitations of integrating a reflection activity into the design process?

Answers to these questions will deepen knowledge about how reflection can promote deeper engagement with feedback and make each iteration productive. Answers will also have implications for design workflows implemented in individual design processes and in feedback platforms.

EXPERIMENTAL DESIGN

To answer these research questions, we conducted a between-subjects online study with five experimental conditions (see Activity Manipulations). All participants created an initial design, engaged in an activity based on their assigned condition, and then revised the design.

All aspects of the study including the number and phrasing of the reflection prompts, design feedback collection, and time allocated to each activity were piloted with three users experienced in design. We revised the study based on their feedback. None of these users participated in the full study.

Participants

We recruited participants online by advertising on design distribution lists in several U.S. academic institutions, and by posting on design-oriented Facebook group pages. We advertised the study as an online competition to create a graphic design for a half marathon. We were contacted by 223 people; and 90 completed the study (60 female). Participants ranged from 18 to 45 years of age (μ=24, SD=6.9). All of them resided in the U.S. Participants completed an initial survey asking about their education and experience in design. They also rated their design expertise on a scale from 1 (novice) to 5 (expert). The average self-rated expertise was 2.1 (SD=1.0), indicating that most participants perceived themselves as novices. This was supported by a majority of the participants (82%) reporting no academic training and minimal experience in design. Each participant was remunerated $30. To motivate effort, we provided five prizes of an additional $20 for the final designs that received the highest quality ratings from three design experts not affiliated with the research team. We refer to the participants in the study as **designers**.

Design Brief

All designers created an initial and a revised design for the following design brief: *You have been hired to design a flyer for a half marathon race called RUN@NYC. The event will be hosted by and held at Central Park in Manhattan, New York City at 7 am on October 1, 2016. Runners can register through the event website www.running-nyc.com. The top three runners will receive a $300 prize each. The goal of your flyer is to encourage participation, be visually appealing, and convey the event details.*

We chose this design brief because we believed that both the designers and the feedback providers would be familiar with the topic. It was also open-ended, yet simple enough to allow for creative solutions in the time allotted. Designers were informed that they could use graphic design software of their choice and could use any online images in the public domain. To ensure the designers constructed their own solution, the use of design templates was prohibited.

Activity Manipulations

Designers performed a planned activity after submitting their initial designs but before revising it. The activities were organized into five experimental conditions:

Reflect-only (R): Designers reflected on the initial design by writing responses to the reflection prompts defined in the study. No external feedback on the design was provided.

Feedback-only (F): Designers reviewed six pieces of external feedback that was generated for the initial design. No reflection exercise was performed.

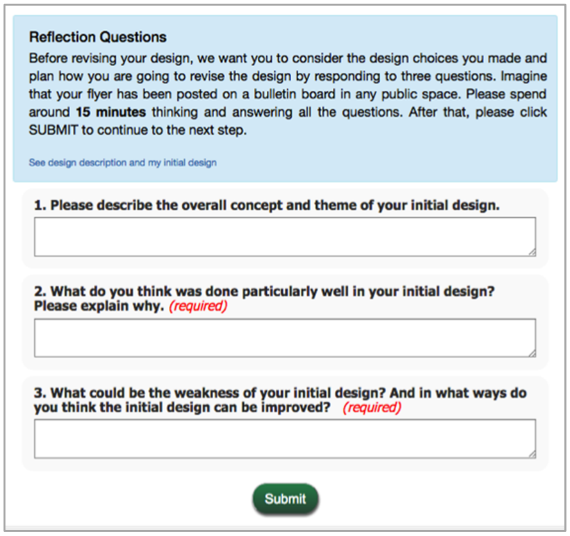

Figure 2. The user interface for a designer to write their responses to the reflection prompts.

Reflect-before-feedback (RF): Designers performed the reflection activity first; then reviewed the external feedback generated for the initial design.

Reflect-after-feedback (FR): Designers reviewed the external feedback generated for the initial design first; then performed the reflection activity.

Control (C): Designers did not perform a reflection activity and did not receive feedback prior to revising the design.

Reflection Prompts

In all of the conditions with a reflection activity (R, RF, FR), designers wrote responses to three prompts (Figure 2):

(1) Please describe the concept of your initial design.

(2) What do you think was done particularly well in your initial design? Please explain why.

(3) What could be the weakness of your initial design? And in what ways do you think the initial design can be improved?

The first question asked the designer to revisit her goals, while the other two prompted an assessment of the design in the context of those goals. The questions were presented two days after the designers submitted their initial design in the associated conditions. The delay allowed the designer time to step outside of their design activity before responding to the reflection prompts. The two-day break was consistent between conditions. The prompts were developed to help designers extract insight from their design activity consistent with Schön's theory of reflection-on-action [40]. The three prompts were derived from a larger pool of possible questions that was refined based on participant feedback in the pilot study.

Feedback Generation

Designers received external feedback on their initial design in the (F, RF, and FR) conditions. A challenge was to deliver feedback that was consistent across designs in terms of length, quality, and scope; yet be specific to each design.

Category	Definition and Example
Theme	**Definition:** Response evaluates the overall direction of the design **Example:** "It doesn't look like a flyer for half marathon race, looks like some flyer of Industrial/business/Real estate."
Surface elements	**Definition:** Response assesses the color, font, and imagery choices made to the design. **Example:** "Your flyer is simple and very good, but the combination of the colors reminds to a party invitation, not marathon."
Composition and layout	**Definition:** Response reacts to the layout and composition of the visual elements in the design. **Example:** "I like the contrast between the top and the bottom. But I do not like the contrast between the words at the top because I'm not sure how visible that will be once you print it out."

Table 1. The rubric used for feedback generation.

To address this challenge, we used a rubric to guide the feedback generation process. The feedback was generated using a paid crowdsourcing platform, Amazon.com's Mechanical Turk (MTurk) [22]. Prior work has shown that the design feedback generated from MTurk is similar in terms of its quality, scope, and length to that generated from social media and online communities [52]. The rubric directed attention to three categories of design feedback: theme, composition and layout, and surface elements. The rubric was adopted from [10], which showed that the use of the rubric increased the quality of crowdsourced feedback to the point of being comparable to expert feedback [54]. Table 1 shows the definition and samples of feedback for each category in the rubric.

On the task page for gathering the feedback, the provider was presented with an image of the design, the design brief, and instructions for providing feedback: *"Your feedback should include both strengths (what you like) and weaknesses (what you don't like) about the [definition of that category of the rubric]. We are not the designer, so you don't need to be overly positive. However, responses that demonstrate insufficient effort or are offensive will be rejected.*" The task was configured to require approval rates at or above 95 percent and prior completion of 500 or more HITs. Five subjects were recruited to provide feedback for each category of the rubric for each design. The research team ranked the quality of the five pieces of feedback, and only the top two pieces of feedback in each category were presented to the designer. Each feedback task paid $0.40, which was consistent with U.S. minimum wage based on pilot data.

When presented with the feedback, the designer rated the perceived quality of each piece of feedback on a Likert scale from 1 (low) to 7 (high). After the experiment, we analyzed the ratings between conditions. An ANOVA did not show a statistical difference between the conditions ($F(2, 373)=0.74$, $p=0.47$) and the feedback quality was reasonably good

(μ=5.18, SD=1.43). We interpret this result to mean that we were successful in generating feedback that was personalized yet was of similar quality and referenced similar themes.

Procedure

After signing up for the study, an account was created for the designer on our review platform. The account information was sent to the designer along with a Web link to access the platform. The study consisted of two phases.

Phase I

In the first phase, the designers were instructed to spend sixty minutes creating their initial design. Once created, the designers uploaded their designs to our platform and completed a survey about their performance (see Measures). A confirmation page was then displayed informing the designers that the second phase of the study would begin in two days. During these two days, the designers did not perform any actions related to the study. However, the research team gathered the design feedback that would be presented in the associated conditions during phase two.

Phase II

After two days, an email was sent to the designers to continue with the study. A page describing the assigned activities was presented to step the designers through the second phase. Designers in the *reflect-only* condition were directed to a page showing the reflection prompts (Figure 2). For the *feedback-only* condition, designers were directed to a page showing two pieces of feedback per category. The designers were instructed to spend 15 minutes performing each activity. For the *reflect-before-feedback* and *reflect-after-feedback* conditions, designers were presented with both activity pages, sequenced according to the assigned condition. Once the activities were complete, the designers were instructed to spend 30 minutes revising their design. Designers in the

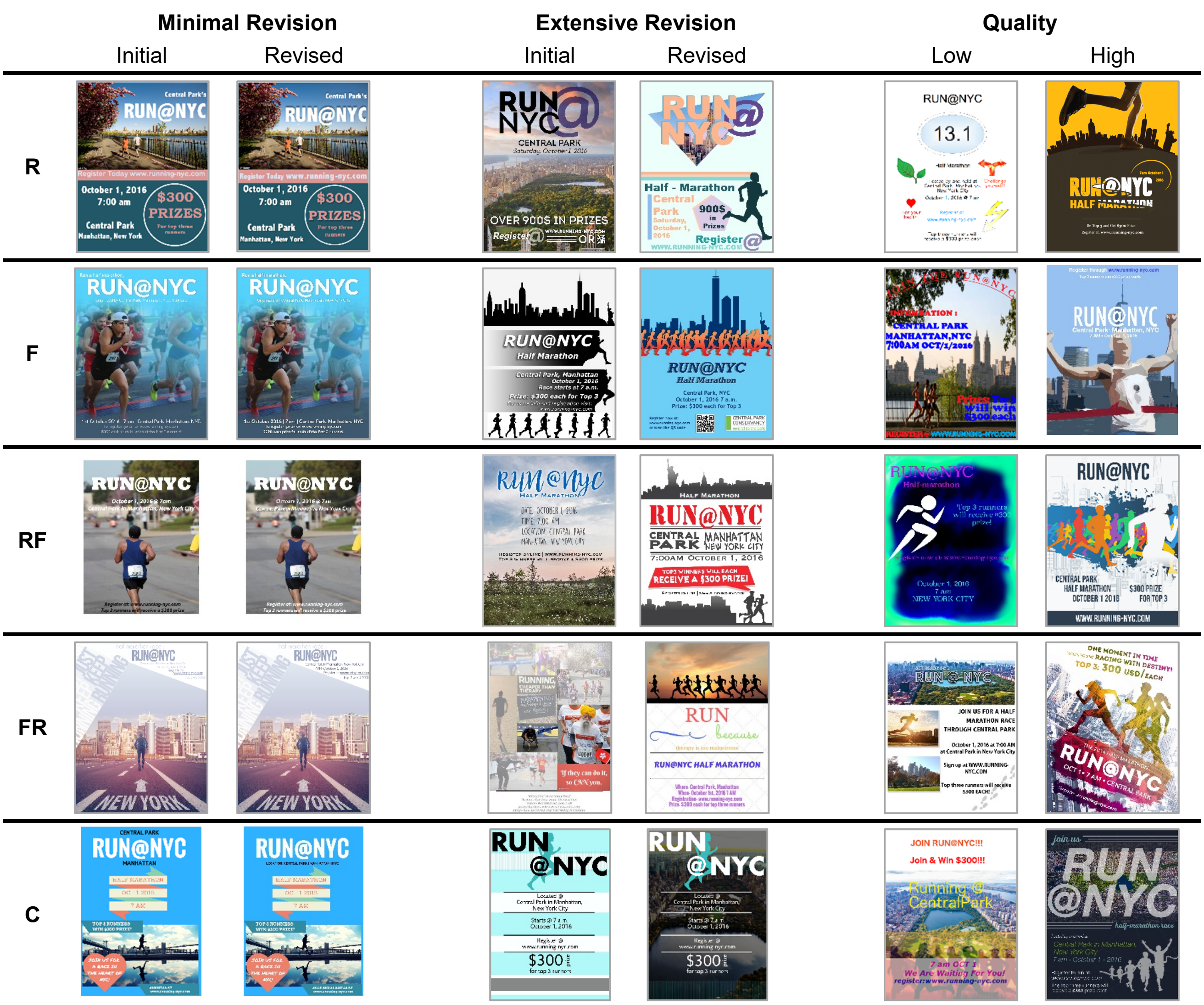

Figure 3. Examples of designs created in each experimental condition. The left two columns show a pair of initial and revised designs where the initial design was minimally revised. The middle two columns show a pair of designs where the initial design was extensively revised. The rightmost two columns show a low and high quality revised design from each condition.

control condition revised their initial design without reflection or reviewing feedback. Designers had up to three days to submit their revised design. All the indicated times were based on results from the pilot study.

A second survey was presented after submitting the revised design. The survey had the same questions as the survey in the first phase but also asked about designers' perceived usefulness of the assigned activities (see Measures).

Measures

In addition to the designs, we collected data through surveys and expert evaluation, and by analyzing the content of the designers' responses to the reflection prompts.

Surveys

At the end of phase one, designers were asked:

(1) How many minutes did you spend creating the design?

(2) How much effort did you invest in the design?

(3) How would you rate the quality of the design?

(4) How confident are you that the flyer fully satisfied the design goals?

At the end of phase two, designers were asked the same four questions along with two additional questions:

(5) Please rate the degree of revision between the initial and revised design.

(6) How useful was [Activity] for improving your design?

Since the control group did not perform any activities, for the last question, they were asked: *"How useful was having two days before being asked to revise their design?"* Questions that asked for ratings were performed on a 7-point scale from 1 (Low) to 7 (High). Designers were asked to briefly explain each rating on the surveys.

Expert Evaluation

Three experts were recruited from the online work platform, Upwork [8], to evaluate the quality of the collected designs. Each expert had eight or more years of professional experience in graphic design and was paid $30. The experts read the design brief and rated how well the design satisfied the goals from 1 (low) to 7 (high). The designs were presented in a random order and the experts were blind to the iteration and condition. The intra-class correlation coefficient (ICC) was 0.83 for the initial designs and 0.80 for the revised designs. These values indicate substantial agreement between the raters to continue with the data analysis [23].

To assess the degree of change between the initial and revised designs, we recruited three additional experts from Upwork, who also have eight years of professional design experience. Additional experts were recruited because the other experts had already been exposed to the designs and we wanted to eliminate potential bias for the ratings of change. The experts read the design brief and rated the 90 pairs of designs. The rating interface randomly placed the initial and revised design side by side. The experts rated the degree of change from 1 (minor change) to 7 (completely different) and were blind to condition. The intra-class correlation coefficient was 0.86, again indicating sufficient agreement between the raters to continue with the data analysis.

Activity	Experts' Ratings		Designers' Ratings	
	Initial	Revised	Initial	Revised
R	3.60 (1.8)	3.75 (1.9)	**4.30 (1.1)**	**5.30 (0.9)***
F	3.26 (1.8)	3.43 (1.9)	4.71 (1.2)	5.07 (1.5)
RF	3.76 (1.9)	3.76 (2.0)	**4.11 (1.2)**	**5.00 (1.1)***
FR	**3.15 (2.2)**	**3.69 (2.0)***	4.67 (1.4)	5.06 (1.3)
C	2.97 (1.6)	3.18 (1.7)	5.15 (1.4)	5.55 (1.1)

Table 2. Mean and (SD) of the quality ratings for the initial and revised designs. The scale is from 1 (low) to 7 (high). An * indicates the ratings of the revised designs were higher than the initial designs in that same condition ($p < 0.05$).

Content Analysis

We categorized the content and measured the length of the reflection responses. Details will be described in the Results.

RESULTS

We collected 90 pairs of initial and revised designs: 20 from the *reflect-only* condition, 14 from *feedback-only*, 18 from *reflect-before-feedback*, 18 from *reflect-after-feedback*, and 20 from the *control* condition. Figure 3 shows examples of designs in each condition.

Design Quality

Table 2 summarizes the experts' and designers' ratings of the perceived quality of the initial and revised designs. The experts rated all of the designs, and the designers rated their own designs. All the ratings were from 1 (Low) to 7 (High).

Experts' Ratings

A paired t-test showed that the revised designs in the FR condition ($\mu=3.69$) had the largest increase in perceived quality relative to the initial designs ($\mu=3.15$; $p=0.001$). For the other conditions, the ratings of the revised designs did not statistically differ from the initial designs. The results show that performing a reflection activity after reviewing external feedback enables the most improvement in design quality according to the expert raters.

An ANOVA also showed that the experts rated the revised designs higher ($\mu=3.56$) than the initial designs ($\mu=3.34$, $F(4, 265)=7.8$, $p=0.005$). This result confirms that iterating on a design also contributes to improved quality [11]. Activity had no effect and there was no interaction effect.

Designers' Ratings

Paired t-tests showed that the designers rated their revised designs in the RF ($\mu=5.00$) and R ($\mu=5.30$) conditions higher than their initial designs ($\mu=4.11$, $\mu=4.30$; respectively, $p<0.05$ in both cases), but not in the other conditions. This pattern was inconsistent with the ratings from the experts (compare the respective data in Table 2). Performing the reflection activity before or without reviewing feedback

Activity	Experts' Ratings	Designers' Ratings
R	2.53 (1.7)	3.9 (1.5)
F	2.57 (1.4)	**4.1 (1.7)***
RF	**2.89 (1.7)***	3.4 (1.6)
FR	**2.67 (1.1)***	**4.7 (1.3)***
C	1.90 (1.1)	2.6 (1.6)

Table 3. Mean and (SD) of the ratings of degree of change between the initial and revised designs. The scale is from 1 (minor change) to 7 (completely different). An * indicates the perceived change was higher in that condition relative to the Control (p<0.05).

resulted in the designers *over-estimating* the increase in the quality of the revised design compared to the expert ratings.

An ANOVA comparing designers' and the experts' ratings also showed that designers rated their designs (μ=4.9) higher than the experts (μ=3.45; F(1,270)=224.5; p<0.0001). One explanation is that the designers in our study reported lower expertise (μ=2.1 on 5-point scale), and prior work has shown that those with less expertise typically over-estimate the quality of their performance [4].

In sum, our results show that performing a reflection activity after reviewing feedback (FR) allows a designer to create a higher quality revised design according to the experts and without over-estimating their own performance.

Degree of Change

Table 3 summarizes the experts' and designers' ratings of the degree of change between the initial and revised designs. The experts rated all pairs of the designs, and the designers rated their own designs. Ratings were on a scale from 1 (minor change) to 7 (completely different).

Experts' Ratings

An ANOVA showed that the Activity had a main effect on the experts' ratings of the degree of change (F(4,270)=2.61, p=0.04). A post-hoc Tukey's HSD test showed that the effect was due to the RF and FR conditions, as the pairs of designs in these conditions were perceived as having more changes than the pairs of designs in the control condition (p<0.05 in both cases). No other differences were detected. Performing a reflection activity either before (RF) or after (FR) reviewing external feedback led to more extensive revisions. Recall that the FR condition also led to the largest increase in perceived quality for the revised designs.

Designers' Ratings

An ANOVA showed that Activity had a main effect on the designers' ratings of the degree of change between their initial and revised designs (F(4, 90)=5.1, p=0.01). A post-hoc test showed that designers in the FR and F conditions rated the degree of change between their designs (FR: μ=4.7, R: μ=4.1) higher than the designers in the control condition (μ=2.6; p<0.05 in both cases). No other effects were found.

An ANOVA also showed that the designers rated the degree of change between their initial and revised designs (μ=3.73) higher than the experts (μ=2.27, F(1, 91)=95.02, p<0.001).

Activity	Confidence		Perceived Effort	
	Initial	Revised	Initial	Revised
R	5.55(0.9)	5.80(1.1)	4.95 (1.1)	5.1 (1.3)
F	5.50 (0.8)	5.79 (1.2)	4.93 (1.1)	4.79 (1.7)
RF	5.50 (0.9)	5.67(1.1)	5.00 (1.3)	4.72 (1.4)
FR	5.32 (1.3)	5.67(1.3)	5.33 (1.1)	5.22 (1.4)
C	5.70 (1.4)	6.15 (0.9)	**5.50 (1.1)***	**4.30 (1.4)***

Table 4. Mean and (SD) of designers' ratings of confidence and effort invested in their designs. The scale is from 1 (low) to 7 (high). An * indicates a significant difference between the revised and initial designs (p<0.05).

The difference in ratings could be due to the designers having been engaged in the design process, and having considered more changes than they actually made and could be observed by the experts in the revised designs.

Designers' Confidence and Perceived Effort

Table 4 summarizes the ratings of the designers' confidence in the initial and revised designs for satisfying the design goals. It also shows the designers' perceived effort invested in the designs. Both ratings were from 1 (low) to 7 (high).

Confidence

An ANOVA showed the designers were more confident in their revised design (μ= 5.8) than their initial design (μ=5.3; F(4, 90)=8.86; p=0.004). No other differences were found. This result shows that the increased confidence was due to iterating on the design [19], and was not affected by performing a reflection activity or reviewing feedback.

Perceived Effort

An ANOVA showed that the designers rated their effort for the initial design (μ=5.16) as higher than for the revised design (μ=4.82, F(1,89)=4.0, p=0.048). This was not surprising, because we asked the designers to spend 60 minutes on the initial design and only 30 minutes on the revised design.

A paired t-test showed that the designers in the control condition reported decreasing the largest effort from the initial (μ=5.5) to the revised design (μ=4.3, p<0.05). Without feedback and without reflection, designers may have lacked direction for the design revisions and exerted less effort.

	RO	FO	RF	FR	C
Reflection	4.6 (1.6)	N/A	4.2 (1.9)	4.7(1.7)	N/A
Feedback	N/A	5.42 (1.3)	6.1 (1.2)	5.6 (1.4)	N/A
Break	N/A	N/A	N/A	N/A	3.9 (1.5)

Table 5. Mean and (SD) of the ratings of the usefulness of performing reflection, reviewing feedback, or taking a break organized by the condition in which the action occurred.

Designers' Perceived Usefulness of the Activities

Designers rated the perceived usefulness of their assigned activity for improving their design. In the FR and RF conditions, the designers rated the usefulness of the reflection activity and reviewing feedback separately. Designers in the control condition rated how useful it was to have a two-day break (i.e. no activity) before revising their

Category	Reflection Example
Overall theme	**Strength:** I think what I did well was reach the target market by persuasion. …The slogan 'this is your year' reinforces the need to sign up. [P27, R] **Weakness:** I feel the poster still lacks a sense of encouragement. I keep thinking about how I can include a motto that is unique, relatable and motivational but not cheesy. [P89, FR]
Surface elements	**Strength:** I think a weakness here is how small some of the font is, particularly in the lower left corner. [P136, R] **Weakness:** I can replace the blue with other colors to make it more visible. The "Top 3 $300 Prizes" text is difficult to see. I can make that more visible by making the outline bolder or changing color. [P90, FR]
Composition and layout	**Strength:** I think I did a good job not making the poster too clustered with words and spacing them out enough so the audience isn't overwhelmed. [P47, FR] **Weakness:** The text was a bit harder to see than it should be. I need to reduce the size of some of the text boxes and edit the background to make it pop a little bit more. [P168, FR]

Table 6. Examples of responses to the reflection prompts.

design. Ratings were from 1 (not useful) to 7 (very useful). Table 5 summarizes the ratings.

Both performing reflection (μ=4.48, SD=1.8) and reviewing feedback (μ=5.6, SD=1.3) were rated moderately useful for improving a design. An ANOVA showed that the condition in which the reflection activity (R, RF and FR) or reviewing feedback (F, FR, RF) were performed did not affect the ratings of their usefulness. The order of performing these actions also did not affect the ratings of their usefulness.

In the RF and FR conditions, designers rated reviewing the feedback more useful (μ=5.81) than performing the reflection activity (μ=4.42, $F(1, 35)$=12.9, p=0.001). Designers in the control condition rated the two-day break least useful (μ=3.9, SD=1.2) for improving their design.

Reflection Responses

We analyzed the content of the designers' responses to the second and third reflection prompts in the R, RF, and FR conditions. These questions related to the strengths and weaknesses of the initial designs. The responses were first partitioned into idea units. An idea unit is a coherent thought captured in a phrase, sentence, or group of sentences [8]. The dataset contained 64, 65, and 64 idea units in the R, RF, and FR conditions, respectively. From an initial pass on the data, we found that the idea units could be labeled using the categories from the feedback rubric: theme, surface elements, and composition and layout. Table 6 provides examples of idea units in each category. Table 7 shows the distribution of idea units by category.

To compare the patterns of interest, we performed z-tests for population proportions, and found that the responses from the RF and R conditions placed more emphasis on the *theme* of the design (RF: 48%; R: 39%) than those from the FR condition (19%; $p<0.05$ in both comparisons). Conversely, the responses from the FR condition had more idea units relating to *surface elements* (43%) than the responses in either the RF or R conditions (RF: 24%; R: 26%; $p<0.05$ in both comparisons). If designers reviewed feedback before engaging in the reflection activity, their attention was directed more toward surface-level issues in their designs such as layout, color, and font choices. The length of the responses (in words) did not differ between the conditions (overall: μ=145 words, SD=18.3).

Category	R	RF	FR
Overall theme	48%	39%	19%
Surface elements	24%	26%	43%
Composition and layout	15%	23%	28%
Other	13%	12%	10%

Table 7. Frequencies of the categories of idea units derived from the reflection responses.

Benefits and Limitations of the Reflection Activity

We analyzed the designers' explanations for their ratings of the usefulness of the reflection activity (R, RF, and FR conditions). The explanations were first partitioned into idea units. The dataset contained 27, 16, and 21 units in the R, RF, and FR conditions, respectively. Our initial pass assigned a theme to each idea unit. Subsequent passes were performed to group similar themes until they were reasonably exclusive. The number of idea units citing a given theme is shown as "(n=<value>)".

When the reflection activity was rated favorably, the most common theme was that it helped the designers revisit their design goals (n=17). One designer stated that he used the reflection activity to better recall his original design ideas:

> *"Writing something out helps me think about what I really mean, helps me clarify my thoughts. This helped me understand what I really thought of the prior design."* [P38, RF]

Designers reported that the reflection activity was also able to help them identify the strengths and weaknesses of their designs (n=14). This was especially true in the R condition (n=11) where designers did not receive external feedback. Intentional reflection enhanced designers' ability to question some of their design choices and inspire revisions. An illustrative quote from a designer was:

> *"At first I wasn't sure how I could improve my design. In fact, I really liked my first design! And then I was thinking of minor things to change, like the font, or the font color etc. But then I thought about the comments that I had made, and honestly felt kind of pressured to make changes based on my actual reflections. After I did make the changes that were consistent with my reflection, I realized how much MORE I liked my new design!"* [P133, R].

A third theme was that designers used the reflection activity to prioritize changes prior to executing them (n=14).

Elaborating on what needed to be fixed promoted a 'stop-and think' gap so that the designers could better consider how to incorporate their ideas into the revised design [24]:

> *"… I had to think through an explanation of my design in the first question, so I was better prepared to answer the second and third questions. While answering questions two and three, I could use the specific points I wrote for question one to decide which elements of the design met the goals and which did not."* [P73, R]

When the reflection activity received less favorable ratings, designers felt that it did not help them generate directions for their revisions (n=10). Such responses came mostly from the R and RF conditions where designers performed the reflection without (or before) receiving feedback:

> *"I don't think they were very memorable or helpful. It didn't really give me any insight on what specific things to change."* [P105, R].

> *"It's always hard to find ways to edit my own work, I feel like it would have been more helpful to receive feedback from someone else and build off that."* [P117, R].

A related theme was that some designers had decided on the revisions based on reviewing the feedback (n=6) and did not find the reflection useful. For example, one designer wrote:

> "*I did not find it useful at all to write about the revisions I was going to make. I had already come up with the changes I wanted to do as I was reading the feedback."* [P90, FR].

Finally, other designers did not value the written forms of the activity, saying that their revisions were already clear in their minds (n=8). One supporting response was:

> *"While the reflection question asks you to think about your design, it is all part of the design process already, so it does not add to improving the design."* [P179, RF].

DISCUSSION AND FUTURE WORK

Our results show a benefit of integrating a lightweight reflection activity within the iterative design process. The feedback-only (F) condition in our study represented the common practice of revising a design based on reviewing feedback from an external audience. Interestingly, we found that performing only the reflection (R) before revising the initial design yielded outcomes that were similar to the feedback-only condition, and either activity led to better outcomes than the control condition. Reviewing feedback helped the designers assess their design through external judgments, while the reflection allowed the designers to extract and articulate what was learned from their experience creating the design [31, 46, 48]. Indeed, designers reported that the reflection activity helped them recall their goals, question their choices, and prioritize revisions.

This finding suggests that when a designer is unable to acquire feedback (e.g. due to the cost of acquiring it), a reflection activity could be performed to guide the next design iteration without losing ground on the measures in our study. However, our results should not be interpreted to mean that a reflection activity could fully replace the need for external feedback in the design process.

If designers perform a reflection activity in addition to reviewing feedback, we found that the degree of revision between the initial and revised design was perceived to be more extensive by the experts. Further, if the reflection was performed *after* reviewing the feedback, the designers improved the quality of the designs the most according to the experts. This finding extends prior work showing that reflection after feedback improves task performance [1] to the context of a creative design task. In our study, the improvement typically came from surface-level changes such as color, font, and layout. Regardless of its placement, integrating the reflection activity did not adversely affect the other measures. It remains an open question whether the focus on aesthetic issues was due to the designers in our study having less experience, or due to the categories and style of the feedback that was provided.

Analysis of the content of the reflection responses indicated a tradeoff between performing the reflection activity before vs. after reviewing external feedback. Designers reported considering more conceptual issues when the reflection activity was performed first; whereas they reported attending to more surface issues when the reflection was performed after reviewing the feedback. It may therefore be most beneficial to perform the reflection activity *before* reviewing feedback during the early design stage when the focus is typically on conceptual issues and to perform it *after* reviewing feedback during the late stage when the focus is typically on improving aesthetics [44]. Future work is needed to confirm these recommendations.

The psychological mechanism for how coupling reflection and feedback works is an open question. If reflection is performed after feedback review (FR), the reflection could be helping the designer to synthesize meaning from the feedback and their design experience [43]. However, it is also possible that the feedback helped direct the designers' attention, thereby making the reflection more effective [1]. There would be an analogous interpretation if performed in the opposite order (RF). Future empirical research is needed to elucidate the mechanism at play and how changing the placement of the reflection activity affects this processing.

Platform developers can draw from a range of mechanisms to aid interpretation of feedback, such as visualizing issue frequency [50], displaying cues of source identity [49], and enabling richer modalities for annotation [53]. A reflection activity could be implemented to complement any of these mechanisms, and have already been implemented in online platforms to promote learning [48]. As implementing the reflection is straightforward, we outline ways to encourage responses. If the reflection is to be performed after feedback review, a platform could request a response before allowing the next iteration to be posted. The response could also be made available to the providers to further ground the feedback exchange [16]. If the reflection is to be performed before reviewing feedback, a platform could release the feedback only after receiving a response. Whether the activity is performed before or after feedback review can be based on the stage of the design or left to the designer.

In our experiment, a designer reflected individually through writing. Alternative formulations of the reflection could include a collaborative dialogue around the prompts [12] or a video recording of one's own responses [18]. Future work is needed to test how different formulations of the activity and different prompts affect design performance.

As in prior design studies [32, 51], our experiment instructed designers to revise their initial designs. It is possible that this procedure limited the scope of changes that the designers were willing to consider. Similar studies in the future may want to instruct designers to create a second new design rather than revising their initial one in order to promote broader design thinking. Designers also wrote their own goals during the reflection activity thereby establishing individual standards. An alternative would have been to show the designers an expert solution in order to establish a shared, external standard. This approach could also contribute to promoting creative effort in the revision.

To check if design expertise affected our results, we calculated an expertise score by adding each designer's self-rated expertise (5-point scale) to the quality of his or her initial design (7-point scale, rated by the expert). With this score added as a covariate, we repeated all of the analyses. The results were nearly identical, indicating that the findings are robust to the range of expertise represented in our participant pool. However, because the range of expertise in our study was skewed toward novices, additional work is needed to test if our results hold for a wider range of expertise.

Two days separated the completion of the initial design and performing the assigned activity in each condition. This time was practically needed to generate the feedback for the associated conditions, though delays also benefit reflection [24]. Future work is needed to test how different durations of this delay would affect the patterns of our results.

Beyond the issues already discussed, we see several exciting directions for future work. First, we want to implement a similar reflection activity in an existing review platform and study how often designers perform the activity and possible incentives, and how it affects their design performance in practice. A second direction for future work is to test the generalizability of our results in other creative domains, such as building architecture, programming, and writing. Third, future work could consider how to re-appropriate the content collected from many reflection activities over time to benefit design learning. Finally, it would be interesting to develop and test reflection activities suitable for different stages of the design process.

LIMITATIONS

One limitation of our study is that it only included two design iterations. It was therefore not possible to test how performing the reflection activity affects an entire design process. A second limitation is that we used a single design brief. This allowed for experimental control but future work is needed to test the generalizability of our findings for different design problems and in different domains. A third limitation is that the designers performed the reflection activity once and this was the likely the first time that they performed this type of activity in the context of solving a design problem. Further studies are needed to understand how the effects of a reflection activity change as experience is gained. Last, the study was conducted online to enable wide participation. Though our study platform stepped the designers through the procedure, it was not possible to verify that they performed the tasks exactly as instructed.

ACKNOWLEDGMENTS

This work was supported in part by NSF awards 1462693, 1530818, 1122206, and 1122320.

CONCLUSION

To promote deeper cognitive processing of feedback received from an external online audience, we studied the effects of combining an explicit reflection activity with feedback review for a creative design task. We found that performing the activity after reviewing feedback enabled the largest improvement in design quality and the most extensive revisions, and without designers over-estimating their own performance. The order in which the reflection activity was performed – before or after reviewing feedback – created a tradeoff between considering more surface-level or more conceptual issues for the design revision. Our results also showed that performing only a reflection activity was as good as only reviewing feedback, and either activity led to better outcomes than performing no activity prior to revising a design. Our work advocates for a future where reflection activities are integrated into creativity tools and individual workflows, as the minimal investment needed for reflection can produce better solutions.

REFERENCES

1. Frederik Anseel, Filip Lievens, and Eveline Schollaert, 2009. Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, *110 (1)*, 23-35. http://dx.doi.org/10.1016/j.obhdp.2009.05.003.
2. Michelle Q. Wang Baldonado and Terry Winograd. SenseMaker: an information-exploration interface supporting the contextual evolution of a user's interests. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM. Atlanta, Georgia, USA. 1997. 11-18. http://dx.doi.org/10.1145/258549.258563.
3. Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. Reviewing reflection: on the use

of reflection in interactive system design. In *Proceedings of the ACM Conference on Designing interactive systems*. ACM. Vancouver, BC, Canada. 2014. 93-102. http://dx.doi.org/10.1145/2598510.2598598.

4. David Boud and Nancy Falchikov, 1989. Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education*, *18 (5)*, 529-549. 10.1007/bf00138746.
5. David Carless, 2006. Differing perceptions in the feedback process. *Studies in Higher Education*, *31 (2)*, 219-233. 10.1080/03075070600572132.
6. Patricia Cartney, 2010. Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35 (5)*, 551-564. 10.1080/02602931003632381.
7. Kate Chanock, 2000. Comments on Essays: Do students understand what tutors write? *Teaching in Higher Education*, *5 (1)*, 95-105. 10.1080/135625100114984.
8. Deanna P. Dannels and Kelly Norris 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. 10.1177/1050651907311923.
9. Jelle Van Dijk, Jirka Van Der Roest, Remko Van Der Lugt, and Kees C.J. Overbeeke. NOOT: a tool for sharing moments of reflection during creative meetings. In *Proceedings of the ACM Conference on Creativity and Cognition*. ACM. Atlanta, Georgia, USA. 2011. 157-164. http://dx.doi.org/10.1145/2069618.2069646.
10. Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer, 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, *17 (4)*, 1-24. 10.1145/1879831.1879836.
11. Steven P. Dow, Kate Heddleston, and Scott R. Klemmer. The efficacy of prototyping under time constraints. In *Proceedings of the ACM Conference on Creativity and Cognition*. ACM. Berkeley, California, USA. 2009. 165-174. http://dx.doi.org/http://dx.doi.org/10.1145/1640233.1640260.
12. G. Fischer, 2000. Symmetry of ignorance, social creativity, and meta-design. *Knowledge-Based Systems*, *13 (7–8)*, 527-537. http://dx.doi.org/10.1016/S0950-7051(00)00065-4.
13. Gerhard Fischer, Kumiyo Nakakoji, and Jonathan Ostwald, 1995. Supporting the Evolution of Design Artifacts With Representations of Context and Intent. In *The ACM Proceddings on Designing Interactive Systems*.
14. Gerhard Fischer, Kumiyo Nakakoji, Jonathan Ostwald, Gerry Stahl, and Tamara Sumner. Embedding computer-based critics in the contexts of design. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. ACM. Amsterdam, The Netherlands. 1993. 157-164. http://dx.doi.org/10.1145/169059.169133.
15. Kristie Fisher, Scott Counts, and Aniket Kittur, 2012. Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* ACM, 247-256. http://dx.doi.org/10.1145/2207676.2207711.
16. Eureka Foong, Steven P. Dow, Brian P. Bailey, and Elizabeth M. Gerber. Online Feedback Exchange: A Framework for Understanding the Socio-Psychological Factors. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 2017.
17. Xun Ge, Ching-Huei Chen, and Kendrick A. Davis, 2005. Scaffolding Novice Instructional Designers' Problem-Solving Processes Using Question Prompts in a Web-Based Learning Environment. *Journal of Educational Computing Research*, *33 (2)*, 219-248. 10.2190/5F6J-HHVF-2U2B-8T3G.
18. Elizabeth Gerber. Tech break up: a research method for understanding people's attachment to their technology. In *Proceedings of the 8th ACM conference on Creativity and cognition*. ACM. Atlanta, Georgia, USA. 2011. 137-146. http://dx.doi.org/10.1145/2069618.2069642.
19. Elizabeth Gerber and Maureen Carroll, 2012. The psychological experience of prototyping. *Design Studies*, *33 (1)*, 64-84. http://dx.doi.org/10.1016/j.destud.2011.06.005.
20. Michael D. Greenberg, Matthew W. Easterday, and Elizabeth M. Gerber. Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers. In *Proceedings of the ACM Conference on Creativity and Cognition*. ACM. Glasgow, United Kingdom. 2015. 235-244. http://dx.doi.org/10.1145/2757226.2757249.
21. Michael D. Greenberg, Julie Hui, and Elizabeth Gerber. Crowdfunding: a resource exchange perspective. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. ACM. Paris, France. 2013. 883-888. http://dx.doi.org/10.1145/2468356.2468514.
22. Joshua Hailpern, Erik Hinterbichler, Caryn Leppert, Damon Cook, and Brian P. Bailey. TEAM STORM: demonstrating an interaction model for working with multiple ideas during creative group work. In

Proceedings of the ACM SIGCHI conference on Creativity and Cognition. ACM. Washington, DC, USA. 2007. 193-202. http://dx.doi.org/10.1145/1254960.1254987.

23. Kevin A Hallgren, 2012. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8 (1)*, 23.
24. John Hattie and Helen Timperley, 2007. The power of feedback. *Review of Educational Research*, *77 (1)*, 81-112. 10.3102/003465430298487.
25. Richard Higgins, Peter Hartley, and Alan Skelton, 2002. The Conscientious Consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, *27 (1)*, 53-64. 10.1080/03075070120099368.
26. Lichan Hong, Ed H. Chi, Raluca Budiu, Peter Pirolli, and Les Nelson. SparTag.us: a low cost tagging system for foraging of web content. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 2008. 65-72. http://dx.doi.org/10.1145/1385569.1385582.
27. Maria Jackson and Leah Marks, 2016. Improving the effectiveness of feedback by use of assessed reflections and withholding of grades. *Assessment & Evaluation in Higher Education*, *41 (4)*, 532-547. 10.1080/02602938.2015.1030588.
28. Wendy Ju, Arna Ionescu, Lawrence Neeley, and Terry Winograd. Where the wild things work: capturing shared physical design workspaces. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM. Chicago, Illinois, USA. 2004. 533-541. http://dx.doi.org/10.1145/1031607.1031696.
29. Scott R. Klemmer, Mark W. Newman, Ryan Farrell, Mark Bilezikjian, and James A. Landay. The designers' outpost: a tangible interface for collaborative web site. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. ACM. Orlando, Florida. 2001. 1-10. http://dx.doi.org/10.1145/502348.502350.
30. Q. Vera Liao and Wai-Tat Fu. Expert voices in echo chambers: Effects of source expertise indicators on exposure to diverse opinions. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 2014. 2745-2754.
31. J. John Loughran, 2002. Effective Reflective Practice: In Search of Meaning in Learning about Teaching. *Journal of Teacher Education*, *53 (1)*, 33-43.
32. Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* ACM. Vancouver, BC, Canada. 2015. 473-485. http://dx.doi.org/10.1145/2675133.2675283.
33. Effie Maclellan, 2001. Assessment for Learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, *26 (4)*, 307-318. 10.1080/02602930120063466.
34. Philip Mendels, Joep Frens, and Kees Overbeeke. Freed: a system for creating multiple views of a digital collection during the design process. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM. Vancouver, BC, Canada. 2011. 1481-1490. http://dx.doi.org/10.1145/1978942.1979160.
35. Kumiyo Nakakoji, Yasuhiro Yamamoto, Shingo Takada, and Brent N. Reeves. Two-dimensional spatial positioning as a means for reflection in design. In *Proceedings of the ACM Conference on Designing Interactive Systems*. ACM. New York City, New York, USA. 2000. 145-154. http://dx.doi.org/10.1145/347642.347697.
36. Les Nelson, Christoph Held, Peter Pirolli, Lichan Hong, Diane Schiano, and Ed H. Chi. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM. Boston, MA, USA. 2009. 1795-1798. http://dx.doi.org/10.1145/1518701.1518977.
37. Margaret Price, Karen Handley, Jill Millar, and Berry O'donovan, 2010. Feedback : all that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, *35 (3)*, 277-289. 10.1080/02602930903541007.
38. Sadhana Puntambekar and Janet L. Kolodner. The design diary: A tool to support students in learning science by design. In *Proceedings of the International Conference of the Learning Sciences* Charlottesville, VA. 1998. 35-41.
39. Donald Schön, 1990. *Reflective Practioner: How Professionals Think in Action*.
40. Donald A. Schon, 1992. Designing as reflective conversation with the materials of a design situation. *Research in Engineering Design*, *3 (3)*, 131-147. 10.1007/bf01580516.
41. Donald A. Schön, *The Reflective Practitioner: How Professionals Think in Action*.
42. Moushumi Sharmin and Brian P. Bailey. ReflectionSpace: an interactive visualization tool for supporting reflection-on-action in design. In *Proceedings of the ACM Conference on Creativity and Cognition*. ACM. Sydney, Australia. 2013. 83-92. http://dx.doi.org/10.1145/2466627.2466645.

43. Giada Di Stefano, Francesca Gino, Gary P. Pisano, and Bradley R. Staats, 2014 (Revised June 2016). Making Experience Count: The Role of Reflection in Individual Learning *Harvard Business School Working Paper*, *14-093* (http://dx.doi.org/10.2139/ssrn.2414478.
44. Maryam Tohidi, William Buxton, Ronald Baecker, and Abigail Sellen. Getting the right design and the design right. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM. Montreal, Quebec, Canada. 2006. 1243-1252. http://dx.doi.org/10.1145/1124772.1124960.
45. Oscar Tomico, Joep W. Frens, and C. J. Overbeeke. Co-reflection: user involvement for highly dynamic design processes. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. ACM. Boston, MA, USA. 2009. 2695-2698. http://dx.doi.org/10.1145/1520340.1520389.
46. Annekatrin Wetzstein and Winfried Hacker, 2004. Reflective Verbalization Improves Solutions: The Effects of Question-based Reflection in Design Problem Solving. *Applied Cognitive Psychology*, *18 (2)*, 145-156. 10.1002/acp.949.
47. W. Willett, J. Heer, and M. Agrawala, 2012. Strategies for crowdsourcing social data analysis. In *Proceedings of the ACM conference on Human Factors in Computing Systems* ACM, 227-236.
48. Joseph Jay Williams, Tania Lombrozo, Anne Hsu, Bernd Huber, and Juho Kim. Revising Learner Misconceptions Without Feedback: Prompting for Reflection on Anomalies. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. Santa Clara, California, USA. 2016. 470-474. http://dx.doi.org/10.1145/2858036.2858361.
49. Y. Wayne Wu and Brian P. Bailey. Novices Who Focused or Experts Who Didn't? How Effort and Expertise Cues Affect Judgments of Crowd Work In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM. San Jose, CA, USA. 2016. http://dx.doi.org/10.1145/2858036.2858330.
50. Anbang Xu, Shih-Wen Huang, and Brian Bailey. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM. Baltimore, Maryland, USA. 2014. 1433-1444. http://dx.doi.org/10.1145/2531602.2531604.
51. Anbang Xu, Huaming Rao, Steven P. Dow, and Brian P. Bailey. A Classroom Study of Using Crowd Feedback in the Iterative Design Process. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* ACM. Vancouver, BC, Canada. 2015. 1637-1648. http://dx.doi.org/http://dx.doi.org/10.1145/2675133.2675140.
52. Yu-Chun Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 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*. ACM. Brisbane, QLD, Australia. 2016. 773-784. http://dx.doi.org/10.1145/2901790.2901820.
53. Dongwook Yoon, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E. Löckenhoff, Steven J. Jackson, Abigail Sellen, and François Guimbretière. RichReview++: Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* ACM. San Francisco, California, USA. 2016. 195-205. http://dx.doi.org/10.1145/2818048.2819951.
54. Alvin Yuan, Kurt Luther, Markus Krause, Sophie Vennix, Steven P. Dow, and Björn Hartmann. 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*. ACM. San Francisco, CA. 2016.