Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers

Michael D. Greenberg
Segal Design Institute
Northwestern University
Evanston, IL USA
mdgreenb@u.northwestern.edu

Matthew W. Easterday
Segal Design Institute
Northwestern University
Evanston, IL USA
easterday@northwestern.edu

Elizabeth M. Gerber
Segal Design Institute
Northwestern University
Evanston, IL USA
egerber@tnorthwestern.edu

ABSTRACT

Feedback is important to the creative process, but not everyone has a personal crowd of individuals they can turn to for high-quality feedback. We introduce and evaluate Critiki, a novel system for gathering design critiques on crowdfunding project pages from paid crowdworkers. Stemming from previous research on crowdfunding project creators and their need for early-stage design feedback, we design and build a working system which fits the need of this population: rapid and inexpensive feedback. To solve issues with critique quality we describe a scaffolding technique designed to assist crowdworkers in writing high-quality critiques. We evaluate Critiki with two field deployments: 1) A randomized controlled experiment with 450 crowdworkers to evaluate the efficacy of the scaffolding technique and 2) A user study with 31 crowdfunding project creators to determine usability and user satisfaction. We contribute to research on Creativity and Cognition by demonstrating a working creativity support system, empirically evaluating the system, and describing how scaffolding approaches can be designed for other crowdsourced tasks.

Author Keywords

Crowdsourcing; Design; Scaffolding; Feedback; Crowdfunding; Mechanical Turk

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Online crowdfunding, the act of asking a crowd of supporters to each contribute a small amount of funds towards a larger monetary goal, has emerged as disruptive new way for creative ventures to start and establish communities of supporters [7, 20]. Since 2008, Kickstarter, the largest of the crowdfunding platforms has raised over $1.1 billion for new creative ventures [2]. In addition to coordinating the collection and transfer of funds, the Kickstarter platform provides a page for each project where the creator communicates the need for funds through a title, video and description text. As such, the way that people communicate their project's goals through this page content has a direct effect on crowdfunding success [17, 13]. However, not every project creator creates a succinct, effective pitch.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
C&C '15, June 22–25, 2015, Glasgow, United Kingdom Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-3598-0/15/06$15.00 DOI: http://dx.doi.org/10.1145/2757226.2757249

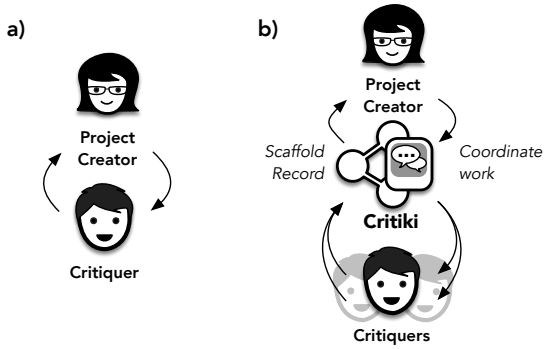

Figure 1. (a) Traditional critique and (b) Critiki: a scaffolded approach to critique. Critiki assists with the process of feedback by acquiring feedbackers, scaffolding the process and presenting results.

Previous research has shown that while the quality of the page content can effect the outcome of the campaign, many crowdfunding project creators do not get feedback on their pages before they launch [15]. This research has shown that the creators of projects are reluctant to ask their personal social network for feedback, are too busy to coordinate a feedback process, and are uncertain how to frame feedback questions [20]. Some project creators don't even have a group of effective feedbackers they can turn to. However, this same research found that people who do muster the time and courage to gather feedback find feedback to be extremely useful in clarifying their pitch and improving their campaign [20].

To address this need, we built Critiki: a system for gathering design critiques on crowdfunding pages from paid crowdworkers (Figure 1). Critiki allows crowdfunding project creators to quickly (<2 days), and inexpensively ($5) gather design feedback with low effort (<5 minutes) on their part. Critiki manages the process of designing effective feedback tasks, posting tasks on paid crowdwork platforms, and collating the feedback so that project creators can use the feedback to revise their project page.

Beyond crowdfunding, the Internet has sparked a generation of individuals who increasingly collaborate on creative work online. As researchers of creativity, we are interested in building new computational systems which support the creative process. In this paper we focus on creative support in the form of feedback, which has been shown to increase the quality of creative work [18, 4]. However, not everyone has access to a crowd of experts who might give high quality feedback. To this end, we're interested in supporting feedback processes, by lowering the barriers to accessing feedback, in domains where people are doing creative work online.

Leveraging paid crowdworkers for feedback however, presents challenges since crowdfunding pages are thematically diverse and have many parts. For example, crowdworkers may not know that the quality of the video on a crowdfunding page is more important than the project description text [17]. In addition crowdworkers might not have the vocabulary to elicit an effective critique. Since the range of themes and possible critiques on crowdfunding pages is diverse, open-response feedback prompts, rather than rubrics or more rigid structures might be an appropriate approach. However, previous research has shown that there are serious issues present with asking crowdworkers open-response questions as well [21].

Critiki addresses issues with open-response questions by doing the following: 1) We use a scaffolding approach from learning sciences to guide crowdworkers through the steps of an effective critique. 2) We provide examples of high-quality critiquing points to crowdworkers to provide them with a vocabulary necessary to elucidate high-quality critiques.

We evaluate Critiki with two field deployments. In the first deployment we show the effectiveness of a scaffolding approach by comparing the quality of scaffolded feedback responses with general open-ended responses in a randomized controlled experiment. In the second deployment, we deploy Critiki with creators of 31 crowdfunding projects and show how they find it useful and easy to use.

This work provides four distinct contributions to Creativity and Cognition: 1) a description of the design and implementation of a system that gathers high quality feedback for crowdfunding project pages, 2) a scaffolding approach for gathering feedback, 3) a set of principles for the design and implementation scaffolded tasks with paid crowdworkers, 4) in the context of crowdfunding we empirically demonstrate that our scaffolding technique is superior to open-response approaches.

RELATED WORK

In this section we review relevant work from crowdfunding, and crowdsourcing as well as related work from the learning sciences on design feedback and scaffolding approaches.

Crowdfunding

Crowdfunding has emerged as a new means for novice creatives to collect funding to do new creative work. Crowdfunding supports creatives by providing a platform to collect funds for and assemble a community around their creative projects in public and in enterprise settings [16, 30]. Previous social computing research on crowdfunding has shown that elements of the campaign page, phrasing of the pitch and updating backers can play a significant role in the success of campaigns [17, 29, 38]. Despite the growing popularity of crowdfunding, few tools have been built that incorporate these findings to assist new project creators with the process of crowdfunding.

In general, design critique and feedback are crucial to all creative success [18, 35]. However, novice creatives, especially those who are running their first crowdfunding campaign often do not have access to experts for critique, nor might they have the time, motivation, or expertise to source high-quality feedback [20]. If we can create a platform to efficiently and inexpensively crowdsource design critique then we can serve this open need. However, crowdsourcing open-response critique presents challenges in achieving complex work.

Open-response Approaches to Crowd Feedback

Significant work has been done in the area of crowdsourcing feedback for creative work. Many commercial systems such as Behance and Dribbble exist for designers to get simple, open-response feedback from an unpaid crowd. Previous research on Dribbble however, has shown that using the site for "showing-off", rather than for gathering feedback is how users attain and grow status in the community [28]. Prior to these online feedback systems, there is a significant body of literature related to automated feedback systems (eg. [14]). While we acknowledge the computational approaches to feedback, in this paper we focus on crowdsourcing as a method of gathering design feedback.

In the area of crowdfunding, Kickstarter has recently implemented an open-response feedback mechanism on their website. However the Kickstarter feedback interface relies on the project creator having an established community of feedbackers s/he could turn to. If project creators eschew the Kickstarter feedback system and attempt to gather feedback off the platform, these creators may not have crowds ready to critique, the time necessary to design critique tasks, nor the expertise to ask effective feedback questions. Asking paid crowdworkers for feedback (as opposed to personal social connections), might solve these issues in gathering design feedback.

Task Design for Crowdwork

An open question in crowdsourcing research is *how to improve crowdwork quality on complex tasks* [22]. The "find-fix-verify" and "shepherding" approaches involve algorithmically dispatching workers to either locate errors or grade the work of other crowdworkers [8, 12]. These approaches work best for tasks which can be easily broken down into independent sub-parts.

While many have attempted to improve the quality of crowdwork through algorithmic approaches, a handful of researchers have examined how the design of tasks affects work quality. Previous research has focused on designing for usability [25], and there is a small but growing body of empirical work on how the design of tasks can affect the quality

of crowdwork output. Research has shown that easier to comprehend and cognitively designed tasks result in better quality work [19, 3].

Scaffolding

One alternative to algorithmic and design based approaches is an approach based on scaffolding: guiding individuals through smaller subtasks in sequence that, in turn, have them complete a larger complex task [32]. Unlike many crowdsourced tasks in which many workers collaborate on one canonical answer, design critique involves independent work. Design critique also involves many steps: observing, knowing what to critique, and expressing the critique in a logical format [26]. Scaffolding, an approach studied by the learning sciences, is widely used across domains such as paper writing, where it has been shown to improve the quality and quantity of student feedback to expert levels [10, 9]. However, challenges remain for crowdsourcing feedback since crowdworkers often lack (1) contextual knowledge, and (2) domain understanding [5]. Drawing from learning sciences research, we use scaffolding to address both of these challenges.

Scaffolding Responses Online

Several projects from Social Computing research have implemented scaffolding approaches with crowds. Kokkalis introduced the idea of taskplans, short checklists of tasks to be done, which helped workers to finish tasks faster and more accurately [23]. Kulkarni and Klemmer take a scaffolding approach to peer grading in MOOCs. In this work they describe how structured peer training and grading mechanisms can perform at the level of expert graders and achieve high levels of agreement with previously established answers [24].

One approach to structuring crowd critique tasks without established answers is Voyant. The Voyant system breaks down the task of critiquing a visual design into several domain-specific sub-tasks, such as determining the focus point, and recording visceral impressions [37]. Voyant is focused on gathering rapid user impressions as a form of critique, and does not focus on text-based critiques. This system works well for visual design tasks, but subtasks in Voyant are inextricably tied to idiosyncratic features of visual design tasks.

Another approach to scaffolding crowd critique is the CrowdCrit system described by Luther [27]. Similar in approach to the Voyant system described above, CrowdCrit is highly specialized to visual critique, with a focus on judging content based on agreed upon principles of high quality graphic design. The process of eliciting critiquing points is supported by the system, but does not support critique of multimedia content such as webpages. In parallel to previous work, our goal with Critiki is to design scaffolding for open-response feedback tasks for crowdworkers where there is no set of established answers. As crowdworkers are potentially novices in giving design feedback [33], we apply scaffolding to guide them through the process of giving high quality design critiques. We hypothesize that scaffolded critiques will be higher quality than open-response critiques (*H1*).

Scaffolding with Examples

While scaffolds can be used to guide workers through the steps of a complex tasks, a challenge remains that paid crowdworkers may not have the domain knowledge necessary to compose a good critique. In other work on crowd critique and peer critique, the peers who are critiquing work have also completed a version of the task for their own submission [10, 24]. Even though these individuals may not be experts in critique, they will most likely be familiar with the assignment or task, as they have previously completed the task at hand. Paid crowdworkers however, do not have this contextual knowledge. Therefore the final challenge in crowdsourcing critique is informing the crowdworkers of relevant contextual information so that they can construct a quality critique.

How then can we quickly educate people about topics for good critique? One approach takes inspiration from how students learn to mirror the behaviors of their teachers [11, 5]. In classrooms it is often observed that language used by teachers is quickly adopted by students [5]. We hypothesize that we can encourage mirroring behavior through the use of examples. Giving crowdworkers example terminology to adopt in each feedback prompt will encourage mirroring behavior and therefore improve the quality of critiques (*H2*). Overall, while crowds have proven to be an effective tool for gathering an anonymous, unbiased opinion on tasks like visual design, and as peer assessors in MOOCs, challenges still remain in crowdsourcing complex, creative work like text-based critique. In this paper we demonstrate how a properly designed scaffolding approach can improve the quality of crowdsourced feedback in the domain of crowdfunding.

Design Rationale

We argue that in order to encourage crowdworkers to give higher quality feedback: (1) the crowdwork platform needs to use scaffolded tasks and guided feedback prompts, (2) crowdworkers need to be briefed on relevant contextual information through examples, and (3) the critiquing system needs to be easy to adapt and extensible, so scaffolded tasks may be iteratively designed and adapted to other domains. Previous research indicates that scaffolded tasks work to produce quality feedback from novices, and that a scaffolding approach can brief individuals on relevant contextual information [10]. The scaffolding approach should solve the challenge of assisting paid, novice, crowdworkers with performing critiques in domains in which they are non-experts. We expect this technique to perform higher quality feedback than open-response prompts.

CRITIKI

Critiki is a scaffolded feedback system powered by crowdwork platforms, such as Amazon Mechanical Turk or MobileWorks. Critiki consists of several interfaces for submitting feedback tasks, performing critiques, and viewing critiques:

Task Creation: the creator of the crowdfunding campaign can submit a crowdfunding project page for critique. See Figure 3.

Task Interface: the crowdworkers view the crowdfunding page to be critiqued, constructs their critique and submits the task. See Figure 4.

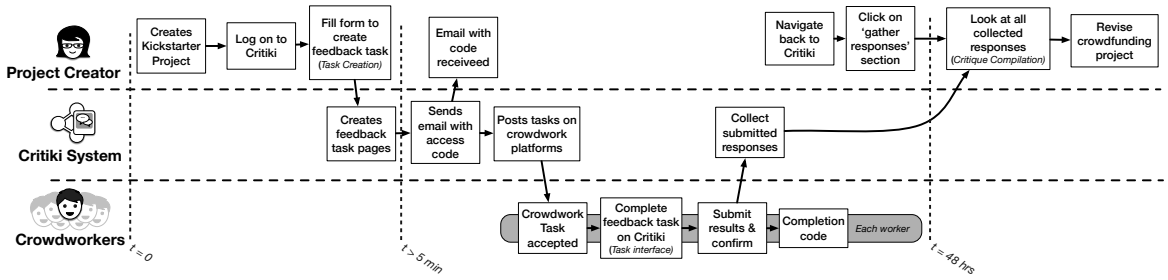

Figure 2. Map of interactions on Critiki

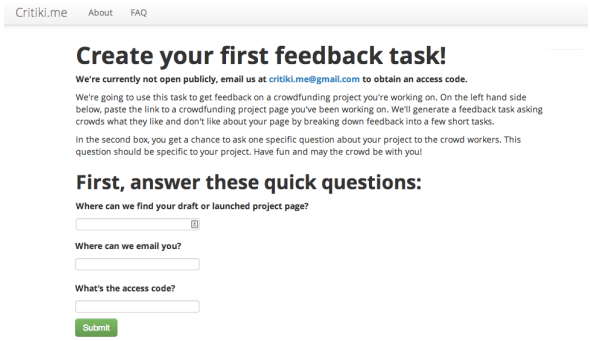

Figure 3. The task creation interface on Critiki. Creators answer three questions to create feedback tasks for their crowdfunding project

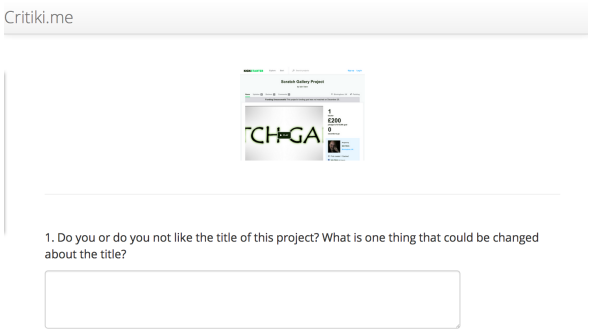

Figure 4. A task on Critiki. The content to be critiqued is on the top. The first two scaffolded questions are at the bottom. Thumbnail can be clicked to open in a new tab.

Critique Compilation: the creator of the campaign can see all the critiques presented on one page.

We use Critiki to gather design critiques on the title and video of crowdfunding campaigns. Critiki was iteratively designed and deployed from April 2013 to January 2014. Critiki facilitates the workflow of the feedback process (Figure 2), from the sourcing of critiques, to the presentation of assembled critiques. This process proceeds as follows:

1. Creator creates campaign content on Kickstarter.
2. Creator submits their crowdfunding page for critique (Figure 3).
3. Feedback Tasks are created on crowdwork platforms.
4. Crowdworkers are recruited on MobileWorks or MTurk [31, 1] and are brought to the task interface (Figure 4).
5. Crowdworkers answer six open-response critique questions, specifically designed to scaffold the critiquing process.
6. Steps 4, 5 and 6 are repeated until all tasks are completed.
7. Creator views feedback. A unique link provides access to a page on Critiki to view aggregated feedback.

Critiki coordinates all of the above steps. A user who is requesting feedback on Critiki needs to provide only three pieces of information to create a task: 1) their email, 2) the location of their project page (a Kickstarter.com URL), and 3) an access code we provide to them (to prevent abuse). Creating a task on Critiki should take under 5 minutes. In contrast, consider the knowledge and time involved in posting a request for work on a crowdwork platform like MTurk. One must first create and fund an account. Next the individual would need to design an unbiased and useful feedback task. One must also know the right amount to pay per task and the proper way to advertise the task on the MTurk. Errors in any of these steps might lead to poor results or no results at all [21]. Platforms like MTurk are certainly valuable resources, but the barrier to entry is often high. With Critiki we aim to lower the barrier to accessing high quality critiques.

In addition, consider the potential social costs for posting feedback tasks on personal social networks. Previous research has shown that people are reluctant to ask their personal social networks for questions that might be personal, and that individuals are reluctant to ask too much of their personal crowds [34]. As such, having an inexpensive tool for gathering high quality feedback quickly could reduce the burden on personal social crowds.

We designed Critiki to be an easy-to-use interface to crowdwork platforms like Mechanical Turk and to avoid the social costs of asking friends and family for feedback. As such, we have both research and usability goals for Critiki. For the user we have designed Critiki to be easy to use, and useful. The research goal for Critiki, however, is to show the efficacy of a scaffolding approach to improve feedback quality.

STUDY DESIGNS

We present two studies with the Critiki system: 1) To empirically evaluate the quality of feedback gathered by Critiki, and 2) To evaluate the usability and usefulness of the system with crowdfunding project creators.

STUDY 1: CRITIKI EVALUATION STUDY

The first study focused on evaluating the quality of feedback between three different feedback task designs. With this study we test our two hypotheses:

H1: Scaffolding tasks will return higher quality critiques than open-response tasks.

H2: Scaffolding with examples will encourage workers to use example terminology.

Study 1: Participants

To evaluate the scaffolds, we recruited workers on MTurk to critique 3 projects. The projects came from the Fashion, Product design and Technology categories on Kickstarter. 150 workers critiqued each project, 450 critiques in total. MTurk workers could critique multiple projects but not allowed to critique one project more than once. MTurk workers were paid $0.25 per response (US workers only). MTurk workers averaged just under 5 minutes on each task. Data collection took place over 10 days. We also recruited two crowdfunding experts to critique each of the three projects. Each expert had launched multiple successful crowdfunding campaigns, raising between $3,000 to over $110,000. Themes of their projects included: poster design, urban gardening, and photography. The experts were thanked for their time with $20 gift cards.

Study 1: Experimental conditions

To determine the efficacy of the scaffolding technique for gathering better critiques, responses on Critiki (step 5 in the above procedure) were framed in one of three ways: Open-Response, Scaffolded, and Scaffolded with Examples. We used a randomized, controlled 3-group between subjects design for this experiment. Below we describe each of the experimental conditions:

Condition 1: Open-Response

In this condition, crowdworkers were asked to give a design critique and were only asked one question. The question asks the crowdworker to "Give feedback on the title and video of the project page." This condition was included as a baseline condition for comparison. This condition is also similar to feedback prompts on popular design feedback platforms online such as Behance and Dribbble as well as the feedback system built into Kickstarter. Our expert responses were also collected in this condition.

Condition 2: Scaffolded

In this case crowdworkers are asked to critique a crowdfunding page by answering 5 questions which scaffold the process of giving a critique. Each question asks the crowdworker to focus on a single element of the campaign and describe what they liked, didn't like and would change about the element. The questions were intended to walk the crowdworker through the process of critique by highlighting important sections of the page. These questions were informed by previous research on page elements which are important to success, [17] and were iteratively designed. These questions focus on the title, video narrative, video length, and video quality. For example, the question about the title asks "Do you or do you not like the title of this project? What is one thing that could be changed about the title?" We hypothesized that scaffolded critique will result in higher quality critiques than open response (H1). We recognize that a common definition of scaffolding includes the concept of "fading," where the scaffolds are gradually removed as skills improve [36]. Our approach does not use fading since we cannot know whether the crowdworker has previously performed tasks. Our approach in this condition is to present the same scaffold to every crowdworker.

Condition 3: Scaffolded with Examples

In the third case we have added a series of examples of high quality critiques to the scaffolded questions. Each question is paired with 4 examples of potential pitfalls which can be critiqued. These examples are also sourced from previous research on effective crowdfunding campaigns and communication design. For example, the question which asks about video quality informs the crowdworkers that: "In good videos: 1) Scenes are well lit, 2) Transitions between scenes are smooth, 3) The volume on voices is even, 4) The camera is not shaky" This condition tests whether crowdworkers will use examples if they are provided. We hypothesize that workers assigned to this case will discuss the themes presented in the hypotheses with a greater frequency (*H1*), and will provide higher quality critiques than the open open-response (*H2*).

Evaluating Critique Quality

To evaluate the quality of the feedback we adopted a technique similar to that used by Cho in previous research on evaluating the quality of scaffolded critique [9]. First, feedback was divided into idea units, which are defined by Cho as "a self-contained message on a single problem" [6], in essence a singular thought or idea. Once the critiques were broken into idea units they were coded into one of 8 categories to separate high-quality critique elements from low-quality critique elements. Table 1 shows examples of idea units and their corresponding codes.

We used a modified version of the Cho and Schunn coding scheme due to differences in our tasks. The Cho and Schunn task is focused heavily on gathering suggestions, while our task is designed to gather suggestions, praise and critique. As such, Cho and Schunn make a distinction between high-quality and-low quality suggestions; they use the terms "directive" and "non-directive" respectively. We use the terms "Nuanced Suggestion" and "Suggestion" to describe the same concept. Since we are looking to evaluate quality we need a distinction between high and low quality praise and critique in addition to suggestions. This coding scheme is designed to flag aspects of positive critiques: 1)Statements telling an individual what to change (Nuanced Suggestions), 2) Statements telling an individual specifically what people like (Nuanced Praise), 3)Statements describing issues people had with specific page elements (Nuanced Critique).

These codes follow 2 of Hattie's "3 Feedback Questions," which must be answered in high quality feedback [18]. Hattie lists these questions as 1) How am I going? (Praise & Critique), and 2) Where am I going? (Suggestions).

A high quality critique will therefore have a higher number of idea units coded as Nuanced Critique, Nuanced Praise

Code	Explanation	Example Responses
NC	Nuanced Critique	"I didn't like the hand motions the speaker was using, it was very distracting." "The main issue is that the volume is waaaay(sp) too high in the beginning"
C	Critique	"The video is low quality." "I don't like the title"
NP	Nuanced Praise	"I really liked that the lighting was bright and you could see the actors." "The video does a good job of telling the audience the point–better quality sneakers."
P	Praise	"The video is good." "Great video."
NS	Nuanced Suggestion	"You should use a pointed microphone to capture better audio" "Generally, I dońt think videos of this sort should last any more than 2 minutes."
S	Suggestion	"The audio should be improved." "Fix the video."
I	Indifference	"The title is OK." "I have no suggestions"
SUM	Summarization	"This is a Kickstarter project." "You are asking people for money"
OT	Other	"Good luck!" "N/A"

Table 1. Critique quality codes

and Nuanced Suggestions. We sum the number of idea units coded as $NC + NP + NS$, and assigned this value as the number of "High Quality Units" (HQ). Conversely, the number of "Low Quality Units" (LQ), was defined as the sum of the remaining codes: $C + P + S + I + SUM + OT$. We define higher quality critiques as having a higher quantity of high quality units than low quality units.

Evaluating the use of Examples

One coder coded each of the idea units for example theme use. Idea units were flagged if they matched the theme of example text. For example, a comment about "lots of camera movement" was coded as a use of the example: "shaky camera". However, if the theme of the comment did not match one of the examples then the idea unit was not coded as such.

Study 1: Results

We used Critiki to gather feedback on three different Kickstarter projects. We gathered 450 critiques (150 per project) from workers on Mechanical Turk. Two crowdfunding experts critiqued each of the three projects for a total of 456 critiques. Each of these critiques was divided into idea units. Two coders who were experts in design critique divided 10% of the data set (46 critiques) into idea units. The coders agreed on their idea units ($ICC(A, 1) = 0.965$), and one coder was tasked with dividing the remaining critiques into idea units. This process resulted in 4447 individual idea units. Each of these idea units were coded by two coders into the eight codes listed above. Two coders again coded a subset of 10% of the data into these eight categories, after two rounds of training the inter-rater reliability for the raters was found to be $\kappa = 0.71$ ($p < 0.001$). One coder coded the remaining 90% of the dataset. Coded data was then summed into *HQ* and *LQ* responses. Table 1 lists the average number of idea units per response that fell into the High Quality and Low Quality categories by condition.

	Average number of Idea Units			
	High Quality	Low Quality	Example	Total
Open Response	3.42	0.94	0.55	4.36
Scaffolded	9.18	3.66	1.57	12.84
Scaffolded w/ Ex.	9.50	3.93	2.44	13.43
Expert 1	6.67	0.67	1.00	7.33
Expert 2	10.00	3.67	2.00	13.67

Table 2. Summary of study 1 results (Average number of idea units per response).

We find that the length (number of characters) of critiques was longer in the scaffolded and scaffolded with examples cases (4.36, 12.84 and 13.43 idea units on average in the open-response, scaffolded and scaffolded with examples cases respectively, see Table 2). Figure 5 shows a box plot of the total number of idea units by conditions. In general we see higher number of idea units in both the scaffolded and scaffolded with examples conditions than in the open-response condition. A pairwise t-test with the Bonferroni correction showed a significant result in each case ($p < 0.001$). There was no significant difference between the scaffolded and scaffolded with examples conditions.

In addition, we find that critiques completed in both the scaffolded conditions produced a higher quantity of idea units coded as High Quality units. This is consistent with *H1*. Fig-

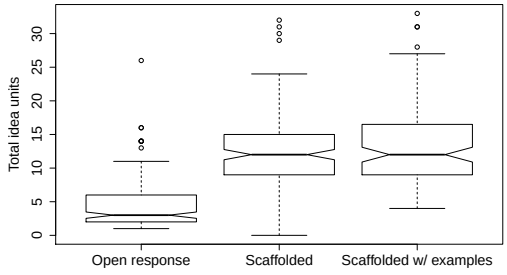

Figure 5. Box plot of the total number of idea units across conditions. Note that both scaffolded cases return more idea units (longer critiques) than the open-response case.

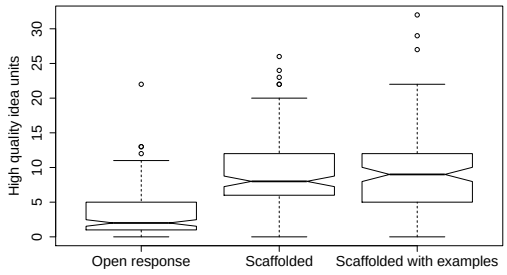

Figure 6. Box plot of number of High Quality idea units across conditions. Note that both scaffolded conditions produce responses with a higher number of high quality idea units.

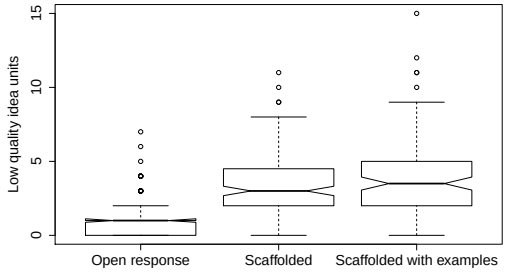

Figure 7. Box plot of number of Low Quality idea units across conditions. Note that both scaffolded cases have a higher number of low quality units.

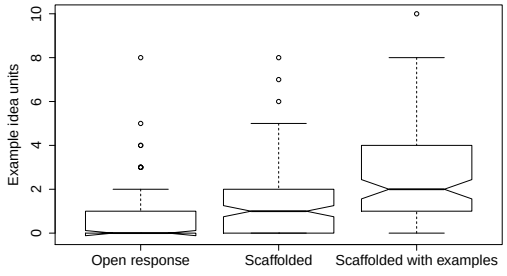

Figure 8. Box plot of number of example idea units across cases. Note increasing trend as we move from open-response to scaffolded and scaffolded w/ example cases.

ure 6 shows a box plot of the High Quality idea units across conditions. A pairwise t-test with the bonferroni correction showed a significant result between cases the open response case and each of the scaffolded cases ($p < 0.001$ in each case). There was no significant difference between the scaffolded and scaffolded with examples conditions.

We also find that both the scaffolded conditions produced a higher quantity of Low Quality idea units as well. Figure 7 shows a box plot of the Low Quality idea units. Pairwise t-tests with the bonferroni correction show a significant difference between the open response and both the scaffolded and scaffolded with examples conditions ($p < 0.001$ in each case). There was no significant difference between the scaffolded and scaffolded with examples conditions.

We also observe people using the examples more in the scaffolded with examples case than in any other condition. This is consistent with *H2*. Figure 8 shows a boxplot of this distribution as well. In general we see an increasing trend, with the open-response condition providing the lowest average number of example idea units (0.55), more in the scaffolded condition (1.55) and the most in the scaffolded with example condition (2.44). Pairwise t-tests with the bonferroni correction show a significant result between each pair of conditions ($p < 0.001$ in each case).

As a test for robustness and generalizability, a one-way ANOVA was used to test for differences in feedback quality between the three projects in high quality units. The number of high quality feedback units did not differ significantly across the three different projects, $F(2, 300) = 0.265$, $p = .768$.

We find mixed results from the two crowdfunding experts. One expert performed roughly at the level of a scaffolded crowdworker, while the other expert performed at a level between crowdworkers in the open-response and scaffolded conditions. Experts were asked to critique in an open-response manner so it is important to note that an expert performed better than the average crowdworker in the open-response condition. Table 2 reports the average number of idea units per response from each of the three conditions as well as the average response of each of the two experts.

STUDY 2: USER STUDY

In this study we evaluate the efficacy and usability of Critiki by testing with actual crowdfunding project creators. We expect that Critiki will be easy to use and will be useful in the crowdfunding process.

Participants

To determine the usability and efficacy of Critiki we tested it with 31 crowdfunding project creators. These creators were recruited by contacting projects off of the Kickstarter "recently launched" page. To be clear, these were projects that had already launched, not creators in the process of designing campaigns. 186 project creators were contacted, and 46

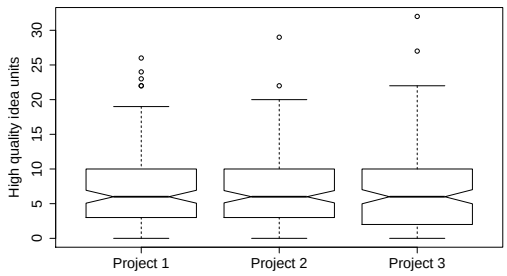

Figure 9. Box plot of number of high quality idea units across projects. Note the lack of variation between different project categories.

individuals agreed to participate; however, only 31 actually followed through created a task on Critiki. Of the 31 creators, 6 were female, 14 were male and 11 projects were run by groups of two or more individuals. Projects came from 8 of the 12 categories on Kickstarter.

For the second study we recruited workers from MobileWorks. For each critique task, 20 individuals were recruited and paid $\$0.25$ (a total of \$5 per project) for a critique. Project creators did not pay to use the Critiki system. However, creators were asked to fill out a short survey about their use and impressions of the system.

Study 2: Data Collection and Analysis

We collected qualitative and quantitative data throughout the month-long deployment. To evaluate the usefulness and efficacy of Critiki, we monitored the quality of the feedback from Critiki and surveyed users for their impressions of the system. The survey asked the users to grade the feedback that was gathered using a rubric for quality, to grade the usability of Critiki, and to grade the usefulness of the system.

Study 2: Results

We facilitated 31 creators of crowdfunding projects to gather feedback on their campaigns using the Critiki tool. 595 individual pieces of feedback were collected; an average of 19.2 responses per project. The full critique was returned to the project creator in around 48 hours.

Feedback Quality

Overall the quality of the critiques spanned from low to high quality, based on an informal appraisal. High quality critiques were often lengthy:

> "It took me a couple times to figure out what it was about. In the description it is trying to be too much. I think it should focus on the story or art, OR focus on the mental health issues addressed but not both and most importantly naming the reviewers on the first page there does not give it credibility but rather it seems to be trying to [sic] hard to prove something."

But other feedback was less specific and actionable, such as general suggestions: *"Show the device detecting bad air quality with cigarette smoke or diesel fumes."*, vague suggestions: *"Vedio [sic] is ok but the description needs some changes."*, and incomprehensible responses: *"ZKZJME"*.

Project Creator Impressions

We distributed a survey to the 31 project creators who tested Critiki with their crowdfunding project, 12 of whom responded (38.1%). Project creators were asked to rate the ease of use for creating and viewing results as well as the helpfulness of Critiki on a 7-point likert scale. $7/12$ reported that Critiki was easy to create tasks on and that Critiki was easy to view feedback with . $10/12$ respondents reported that they would use Critiki again, and $6/12$ reported that they would recommend it to a friend.

Project creators were also asked to rate the usefulness, specificity and comprehensibility of Critiki feedback on 5-point likert scales. We also asked for open-ended thoughts about the Critiki system. Most of the respondents reported that the feedback was useful ($7/12$), specific ($8/12$), and comprehensible ($10/12$). Most respondents felt that Critiki should be used earlier in the crowdfunding process ($8/12$). When asked, one project creator wrote:

> "I didn't use it until after my campaign had launched, though, so my ability to make changes was low." (P1)

When asked about the helpfulness of Critiki in creating campaign material, respondents cited the value of outside opinions, and the objectivity of critiques as benefits. Several respondents did claim that irrelevant and low-quality responses were distracting. One respondent did not understand that Critiki was not going to provide financial contributions to his campaign. However several project creators highlighted the strengths of crowdsourcing feedback. One person wrote:

> "It offered me the perspective of the client as opposed to the perspective that is inherently mine. It was a good way to see what other see, honestly and with helpful ideas ... The person that helped me was great. He gave me some incredible feedback and I used several of his ideas." (P4)

Our small field deployment validated the need for the Critiki system by deploying with actual crowdfunding project creators. While our survey response sample size is small and potentially skewed by the "good subject effect," we found that crowdfunding project creators generally viewed Critiki as useful. In the future, testing with crowdfunders who are *preparing* campaigns, rather than individuals who have already launched campaigns will better validate whether Critiki helps project creators to succeed at crowdfunding.

DISCUSSION

Across two field deployments we evaluated Critiki with both actual crowdfunding project creators and in a randomized controlled experiment. In the first study we show how a properly designed scaffold can be applied to elicit a higher quantity of quality critiques than comparable open-response critiques (validated *H1*), despite there also being a higher quantity of low-quality critiques as well. We feel confident that while scaffolding produces more critiques (both high and low quality), that the benefits of more high quality critique outweigh the potentially negative experience of sifting through more low-quality critique. In our coding scheme the low quality critiques aren't necessarily bad advice, but rather are just

phrases to be skipped over before reaching specific feedback. In addition, we discuss later in this section some potential future approaches to removing low-quality feedback.

We also find that crowdworkers use example terminology when given in the question prompt (validated *H2*). We contribute to the crowdsourcing literature by describing the design and implementation of an effective system and demonstrating empirically how the design of tasks can lead to higher quality critiques. We see this work as a parallel thread to existing, algorithmic, crowdwork approaches [21, 22].

While we showed in the first study that scaffolded crowdworkers perform roughly at the level of crowdfunding experts, we find several other benefits of asking an anonymous crowd of workers for feedback. As we know from previous research, one potential benefit of asking a crowd of individuals for feedback is the ability to gather a wide range of responses [21]. We see this benefit in our study as well. Based on an informal evaluation, we find that all of the themes which experts brought up in their critiques existed somewhere within the 50 responses from MTurk workers. While not every critique discussed every theme, the crowdworkers more than spanned the thematic space that the experts discussed. In addition, crowdworkers bring a diverse set of opinions and provide more opportunities to find smaller mistakes. For example, in one of the projects one crowdworker discussed how the hand gestures the presenter was using were "incredibly distracting". While this is an obscure and possibly minority opinion, this critique might not have surfaced unless many individuals were consulted.

Another benefit of asking a crowd (as opposed to an expert) is the ability to quickly gather consensus [21]. While experts can have gravity attached to their judgments, the same can be given to many individuals saying the same thing. For example, in one of the projects, 22 out of the 50 crowdworkers mentioned that the lighting in the video was too dark. When reviewing all the feedback, this mistake emerges as a common theme. While we did not analyze the feedback for theme consensus within our feedback, we suggest that an interface which presents common critiques (rather than Critiki's current approach of presenting all the feedback, unfiltered) could be an interesting addition to the Critiki design.

In the second study we validated both the need and the efficacy of Critiki by deploying with 31 crowdfunding project creators. While we were not able to recruit people who were actively developing crowdfunding projects, we were able to test with people who had recently launched their project. However, this allowed us to further validate the need for a tool like Critiki in the early stages of crowdfunding planning, as several of the people we recruited claimed that Critiki would have been more useful earlier on in the campaign process. Interestingly, one of the experts we recruited actually inquired about using Critiki to gather feedback on his next crowdfunding project.

Scaffold Design

Due to the malleability of scaffolds and the success of scaffolds in many domains within the learning sciences [10], we feel that scaffolding techniques can be used to used to crowdsource critique in other online domains such as website design and presentation design. However, designing a quality scaffold is not quite as simple as asking a few general questions. Through the design process for Critiki we discovered that our best scaffolded questions for feedback asked for specific contributions as well as specific changes to be made.

When designing an effective scaffold it is also important to first understand the limitations of the system to be critiqued and potential pitfalls in critique. For example, in early iterations of the Critiki system, our scaffolds did not specify exactly what segment of the page was up for critique. This led to some critique opinions which might be valid, but not high quality suggestions. For example, some early critiques we gathered critiqued the "green color scheme" on the page. Unfortunately, on a Kickstarter page, this is not an editable element. To this end, scaffolds must be designed to guide the workers towards giving suggestions that are both relevant to the page and relevant to constraints of the design space.

Testing scaffolding with the crowd gives an opportunity to iteratively test and alter their scaffold. This is in contrast to deploying scaffolds with students in classrooms, where there is effectively only one opportunity to implement the scaffold correctly (the iteration cycle is at the semester or year level). However, with crowds we can rapidly test scaffolds to determine which parts are effective and which parts are not. This process was crucial to the success of the scaffolds we deployed with Critiki.

Limitations & Future Work

One limitation in our approach is data presentation. We currently present all the feedback that Critiki gathers grouped by question. This interface presents the answers to each question in a bulleted list, as opposed to a narrative or a thematic consensus. One major focus of additional work in this area will be to learn how to automatically separate high-quality critiques from low-quality ones, and to aggregate high quality critiques by theme. This might be an area where a crowdsourcing approach like find-fix-verify, or an algorithmic approach like natural language processing might help to separate and curate critiques into a narrative or checklist.

Future work will focus on implementing an iterative design for the overall task. Currently we ask all 20 crowdworkers to answer the same set of scaffolded questions. We do not know at what point there exist diminishing returns due to the potential thematic repetition in answers. If we understand the point of diminishing returns then it might be possible to design a series of different scaffolded tasks to fill the 20 responses, rather than one task for all 20.

CONCLUSION

We built and evaluated Critiki, a system for gathering design feedback for crowdfunding project creators. This research suggests that with proper design a crowdsourcing system can be built to provide a useful service (design feedback), to a population in need (crowdfunders), for a small amount of money ($5.00). Critiki lowers the bar to accessing high-quality critique, and in doing so it has the potential to help novice designers get critiques before they develop a personal

network of high-quality critiquers. We look forward to developing and deploying more Critiki-like tools and to continuously make it easier to access high quality critiques.

ACKNOWLEDGMENTS

We would like to thank all members of the Delta Lab for their helpful feedback through the process of this project. Special thanks to Darren Gergle and Haoqi Zhang for their detailed critiques along the way. This work was supported by the Segal Design Institute, NSF GFRP and NSF Grants IIS-1320693 & IIS-1217225.

REFERENCES

1. Amazon mechanical turk - welcome. https://www.mturk.com/mturk/welcome.
2. Stats kickstarter. http://www.kickstarter.com/help/stats.
3. Alagarai Sampath, H., Rajeshuni, R., and Indurkhya, B. Cognitively inspired task design to improve user performance on crowdsourcing platforms. CHI '14, ACM (New York, NY, USA, 2014), 3665–3674.
4. Amabile, T. M., Collins, M., Conti, R., Phillips, E., Picariello, M., Ruscio, J., and Whitney, D. Creativity in context: Update to the psychology of creativity, 1996.
5. Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., and Norman, M. K. *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons, 2010.
6. Artemeva, N., and Logie, S. Introducing engineering students to intellectual teamwork: The teaching and practice of peer feedback in the professional communication classroom. *Language and Learning across the Disciplines 6*, 1 (2002), 62–85.
7. Belleflamme, P., Lambert, T., and Schwienbacher, A. Crowdfunding: Tapping the right crowd. *Journal of Business Venturing* (2013).
8. Bernstein, M. S., Brandt, J., Miller, R. C., and Karger, D. R. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In *UIST 2011* (2011), 3342.
9. Cho, K. Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. 260–294.
10. Cho, K., and Schunn, C. D. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. 409–426.
11. Clark, R. Design document for a guided experiential learning course. *Final report on contract DAAD* (2004), 19–99.
12. Dow, S., Kulkarni, A., Klemmer, S., and Hartmann, B. Shepherding the crowd yields better work. CSCW '12, ACM (New York, NY, USA, 2012), 10131022.
13. Etter, V., Grossglauser, M., and Thiran, P. Launch hard or go home!: predicting the success of kickstarter campaigns. In *Proceedings of the first ACM conference on Online social networks*, ACM (2013), 177–182.
14. Fischer, G., Nakakoji, K., Ostwald, J., Stahl, G., and Sumner, T. 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 (1993), 157–164.
15. Gerber, E. M., and Hui, J. Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction (TOCHI) 20*, 6 (2013), 34.
16. Gerber, E. M., Muller, M., Wash, R., Irani, L. C., Williams, A., and Churchill, E. F. Crowdfunding: an emerging field of research. In *CHI 2014*, ACM (2014), 1093–1098.
17. Greenberg, M. D., Pardo, B., Hariharan, K., and Gerber, E. Crowdfunding support tools: predicting success & failure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (2013), 18151820.
18. Hattie, J., and Timperley, H. The power of feedback. *Review of educational research 77*, 1 (2007), 81–112.
19. Huang, S.-W., and Fu, W.-T. Systematic analysis of output agreement games: Effects of gaming environment, social interaction, and feedback. 61801.
20. Hui, J. S., Greenberg, M. D., and Gerber, E. M. Understanding the role of community in crowdfunding work. CSCW (2014).
21. Kittur, A., Chi, E. H., and Suh, B. Crowdsourcing user studies with mechanical turk. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (2008), 453456.
22. Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work* (2013), 13011318.
23. Kokkalis, N., Huebner, J., Diamond, S., Becker, D., Chang, M., Lee, M., Schulze, F., Koehn, T., and Klemmer, S. R. Automatically providing action plans helps people complete tasks. In *Proceedings of the 4th Human Computation Workshop, HCOMP*, vol. 12 (2012).
24. Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., and Klemmer, S. R. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI) 20*, 6 (2013), 33.
25. Lasecki, W., Miller, C., Sadilek, A., Abumoussa, A., Borrello, D., Kushalnagar, R., and Bigham, J. Real-time captioning by groups of non-experts. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, ACM (2012), 23–34.
26. Lawson, B. *How designers think: the design process demystified*. Routledge, 2006.
27. Luther, K., Pavel, A., Wu, W., Tolentino, J.-l., Agrawala, M., Hartmann, B., and Dow, S. P. Crowdcrit: crowdsourcing and aggregating visual design critique. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM (2014), 21–24.
28. Marlow, J., and Dabbish, L. From rookie to all-star: Professional development in a graphic design social networking site. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, ACM (New York, NY, USA, 2014), 922–933.
29. Mitra, T., and Gilbert, E. The language that gets people to give: Phrases that predict success on kickstarter. CSCW (2014).
30. Muller, M., Geyer, W., Soule, T., and Steven, D. Grassroots innovation and collaboration through enterprise crowdfunding. In *Seventh International AAAI Conference on Weblogs and Social Media* (2013).
31. Narula, P., Gutheim, P., Rolnitzky, D., Kulkarni, A., and Hartmann, B. MobileWorks: a mobile crowdsourcing platform for workers at the bottom of the pyramid.
32. Reiser, B. J. Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences 13*, 3 (2004), 273–304.
33. Ross, J., Irani, L., Silberman, M., Zaldivar, A., and Tomlinson, B. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, ACM (2010), 2863–2872.
34. Rzeszotarski, J. M., and Morris, M. R. Estimating the social costs of friendsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 2735–2744.
35. Shute, V. J. Focus on formative feedback. *Review of educational research 78*, 1 (2008), 153–189.
36. Van Merriënboer, J. J., and Kirschner, P. A. *Ten steps to complex learning: A systematic approach to four-component instructional design*. Routledge, 2012.
37. Xu, A., Huang, S., and Bailey, B. Voyant: Generating structured feedback on visual designs using a crowd of non-experts. In *Proc. CSCW, 2014* (2014).
38. Xu, A., Yang, X., Rao, H., Fu, W.-T., Huang, S.-W., and Bailey, B. P. Show me the money! an analysis of project updates during crowdfunding campaigns.