

Enabling Expert Critique with Chatbots and Micro Guidance

Carlos Toxtli, HCI Lab, West Virginia University

Joel Chan, University of Maryland

Walter S. Lasecki, University of Michigan

Saiph Savage, HCI Lab, West Virginia University and Universidad Nacional Autonoma de Mexico

1. INTRODUCTION

Feedback is essential to creative work. Creators can receive many kinds of feedback to their work, from informal reactions/kudos, to more detailed critical analyses that can significantly improve outcomes [Dannels and Martin 2008]. Experts have historically provided critique to creative work within physical studios by being directly collocated with creators, whom they had never met before. However, getting experts critics and creators together at the same time in one physical space is hard. Experts generally have limited time, complex schedules, and are distributed across the world. To overcome these difficulties, online platforms have emerged to support and act as a companion to physical studios [Dannels and Martin 2008]. Such systems, however, assume that experts and learners have met previously offline at a design studio, and consequently fail at connecting individuals who have never physically met. As a result, such platforms usually have a limited number of experts. There are however, many other tools that do facilitate interactions between experts and learners who have never met offline, e.g., online forums like reddit.

Here, learners can post photos/video of their creative work; and then their peers or experts provide feedback to the creative artifacts. However, experts on online forums generally gets stuck on understanding what the creator tried to make [Kou and Gray 2017]. As a result, expert critique is rare [Luther and Bruckman 2008]. Another problem is that experts usually have concerns about providing feedback on forums [Dannels and Martin 2008] and thus prefer to not participate in the activity due to fears of saying something wrong and damaging their reputation [Xu and Bailey 2012]. Given the difficulties of coordinating specialists online, recent research [Xu et al. 2015] has focused on obtaining critique from non-experts, e.g., crowd-workers. However, individuals also use critique to learn about best practices, new topics, and even to network [Dannels and Martin 2008]. Expert critique is therefore still needed, and should be something that researchers aim to facilitate, especially at scale to benefit and empower more learners. To enable creators access to expert critique at scale, we



Fig. 1. Overview of MATT's workflow: 1.- Learner submits to MATT her creative work. 2.- MATT finds an expert, sends the work to the expert, who is guided to review and provide micro-feedback approximating critique about the work. 3.- MATT then presents the micro-feedback from the expert to the learner who can use it to improve her work.

introduce MATT, a chatbot that guides crowds of experts to critique creative work, especially that of novices, early in the design process.

2. MATT: ENABLING CRITIQUE FROM EXPERT CROWDS

MATT consists of two main components: 1) the “Learner Helper” that collects learners’ creative work, distributes the work to experts, and then shares experts’ feedback to learners; and 2) an “Expert Micro-Guidance” tool that orchestrates experts to volunteer quality micro-feedback – which resembles online critique – to help learners at scale.

The Learner Helper component has three functions: 1) to allow learners to easily submit their creative work; 2) to find relevant experts who can critique their work; 3) to present back to the learner the feedback from experts. MATT’s Expert Micro-Guidance module focuses on orchestrating experts to produce, in short bursts of time, quality feedback that resembles critique. The module has 4 features to enable this interaction: 1) *Critique in Short Bursts of Time*. Experts’ time is limited and they also generally lack knowledge of how to effectively produce online critiques. MATT tackles this problem by guiding experts to provide critique to creative work by leveraging task decomposition from crowdsourcing. 2) *Critique Anywhere*. MATT communicates via Facebook Messenger with experts. This design facilitates portability and on-the-go experiences as experts can provide feedback wherever they use Facebook messenger. We believe that these two functions enable more experts to participate in online critique, as they no longer have to invest consecutive hours at a physical desk reviewing work. 3) *Privacy*. MATT integrates into its Experts’ Micro-Guidance module the Learner Helper module to provide privacy to experts. Our design builds upon privacy research that showcases that with anonymity higher quality feedback is produced. 4) *Conversational*. MATT guides experts to produce critique within a conversational setting through chatbots. The conversational aspect of MATT might also help experts to not feel that MATT’s guidance is too dictatorial. Fig 1 presents an overview of this workflow.

3. EVALUATION: FIELD EXPERIMENT

Our evaluation asks: Do chatbots micro-guiding experts enable a better approximation of the gold standard of studio design feedback? To answer this question, we conducted a field experiment to compare the feedback experts generated on MATT to alternative interfaces.

3.1 Method.

We conducted a between subjects study, where participants were divided into 3 conditions that studied different guidance and mediated communication settings: 1) chatbot lacking micro-guidance; 2) chatbot with micro-guidance, MATT; 3) online forum with guidance. We studied the condition of “online forum with guidance” as online forums are a mediated communication channel that experts use frequently. Previous work has already identified that forums without guidance rarely lead to critique as experts spend most of their time interpreting the work [Kou and Gray 2017].

For all conditions, experts provided feedback to the finished creative work of learners. Each expert provided feedback on the designs from different learners. We compared the type of feedback that experts produced across each condition by using qualitative coding to categorize experts’ feedback into the categories developed by Dannels and Martin [Dannels and Martin 2008] in their feedback typology. After providing the feedback, experts completed a survey on their perceptions of the interface they just used.

3.1.1 Categorizing Experts’ Feedback.. To categorize experts’ feedback and study how much critique experts produced, we recruited 2 college-educated Upworkers and asked them to categorize each piece

of feedback into either: “*reactive*”, “*direction*”, or “*critique*”, which are the categories in the feedback typology of Dannels and Martin [Dannels and Martin 2008].

Reactive Category: emotional or visceral feedback that does not provide information on how to improve the work. Examples: “That’s wonderful! Great work!” or “Horrible!”

Direction Category: In this form of feedback, the individual providing the feedback tries to bring the design more in line with her own expectations of what the solution should be. The feedback provides direction but no reasoning behind it. Examples: “I would have...” or “I wish...”

Critique Category: This type of feedback focuses on identifying decisions made in the creative work; relating that decision to an objective or best practice; and then describing how and why the decision made supports or does not support the best practices [Luther et al. 2015].

Participants We recruited learners and experts primarily using social media. Learners were recruited from randomly selected Facebook groups related to learning design and experts were people who on LinkedIn stated they worked in design related areas. All experts volunteered their time. We recruited 548 learners, and 76 experts. Learners submitted 153 creative work pieces to the MATT condition, 213 to the online forum, and 128 to the chatbot without guidance conditions. Learners submitted real world creative work pieces that were from one of these types: website design; poster design for an NGO; t-shirt design for an organization, and an entrepreneur product design.

Results: Types of Expert Feedback The two coders classified each of the 548 messages into the category that represented the message the most (either critique, reactive, or direction). The two coders agreed on the classification of 90.1% of all the feedback produced by experts (Cohen’s kappa = .89: Strong agreement). We then asked a third coder to act as a tiebreaker in cases of disagreement.

Given that we are primarily interested in whether MATT increases the amount of critique that a learner receives, we conducted a logistic regression predicting the likelihood that a piece of feedback would be classified as critique given its source. The logistic regression model showed that a piece of feedback was significantly more likely to be classified as critique when it came from MATT, compared to feedback from the online forum condition ($B=1.12$, $z=4.96$, $p < .01$) or the Bot No Guidance condition ($B=0.83$, $z=3.31$, $p < .01$). The overall model was a statistically significant fit to the data, Likelihood Ratio Test $\chi^2(2) = 26.33$, $p < .01$.

Experts stated they enjoyed moderately the chatbot interfaces (mean=4.85 for MATT and for the chatbot without guidance). The forum interface was also enjoyed, but slightly less (mean=4.77). Experts considered all interfaces to be moderately easy to use (mean=4.8). Open ended responses reinforced that experts felt that MATT helped them to produce meaningful feedback by directing the communication into what mattered: “...Chatbots can direct communication efficiently which you don’t really get with other technology [...]” Some experts expressed that the automated aspect of MATT made its guidance not feel imposing because there was nothing personal about it. It was “just” a machine: “Machines don’t have feeling at all, so also nothing to feel on my side.” MATT’s automation also seemed to help experts accept its guidance, as they felt that machines were made to help humans in their daily work. Experts felt that MATT addressed their privacy concerns (median=5).

Unlike face-to face, MATT has delays in its interactions. Experts never responded immediately to MATT’s requests to provide feedback, but rather responded whenever they were free.

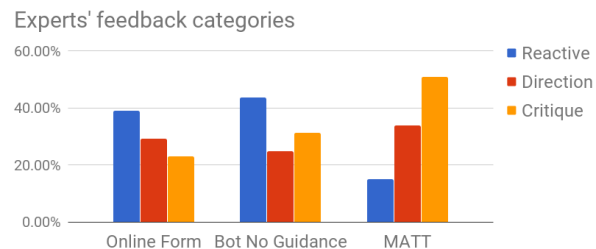


Fig. 2. **Overview of the type of feedback experts generated. Overall, experts using MATT generated the most critiques.**

REFERENCES

- 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 (2008), 135–159.
- Yubo Kou and Colin Gray. 2017. Supporting Distributed Critique through Interpretation and Sense-Making in an Online Creative Community. (2017).
- Kurt Luther and Amy Bruckman. 2008. Leadership in online creative collaboration. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 343–352.
- Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2015. Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 473–485.
- Anbang Xu and Brian Bailey. 2012. What do you think?: a case study of benefit, expectation, and interaction in a large online critique community. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 295–304.
- Anbang Xu, Huaming Rao, Steven P Dow, and Brian P Bailey. 2015. A classroom study of using crowd feedback in the iterative design process. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. ACM, 1637–1648.