# CRITIQUING THE CRITIQUE: THE ROLES OF THE GROUP AND THE EXPERT IN FEEDBACK DELIVERY

Aaron Brown, Quentin Perrot, Ishita Prasad
{abrown94, qperrot, ishitap}@stanford.edu

## ABSTRACT

Creative critique is a complex social and educational tool that allows designers to iterate on their designs at key times during a project's realization. Different methods of delivery can introduce both positive and negative forces to a critique, influencing the general outcome of a student project. We present a study that investigates how the roles of the *expert* and the *group* in varying methods of delivery influences the success of web design critiques. Specifically, we examine how the role of group and expert contribute to answering student designers' questions, offering concrete improvements, and influencing post-critique confidence levels. While our results were not statistically significant, we identify underlying trends that contribute to the effectiveness of a critique, including different kinds of idea generation in critique styles, reduced student participation in the presence of an expert, and the influence of an expert's pre-existing critique style.

## INTRODUCTION

Creative critique, or feedback, in design education is critical to the positive outcome of a student project and thus to the achievement of a student's learning goals. Taylor and McCormack [1] claim that the objective of design critique is to construct knowledge about a design project. More concretely, critiques allow student designers to articulate their goals and assumptions behind their work and recognize how others perceive their work [2]. This process allows designers to identify gaps between intention and interpretation. As a learning designer, it is therefore imperative to receive feedback often to be able to quickly iterate toward an effective solution at each stage of the design process -- allowing the incorporation of suggested elements that arose during critique and the rethinking of more problematic elements.

Critique styles used for visual design, and more precisely, for web design, are of particular interest because their relevance extends beyond the classroom and into industry. Techniques developed in the classroom can be adapted to industry, where the goals of critique are similar. Moreover, there is a growing emphasis in undergraduate education on vocational learning, and as such, further investigation into preferable critique styles can prepare students for professional destinations. The established critique-practice in design schools is face-to-face project-based assessment. However, critique in the creative space of visual design is under-researched [3] and there has been little inquiry into how different forms of critique can improve general outcomes.

Many aspects of the critique-norm, face-to-face critique, are problematic. Judging creative elements is an emotionally-charged event, work is often at a developmental stage and has not been viewed before, and feedback should be balanced in positivity [1]. It is not surprising, then, that critique is described as being "one of the most difficult, demanding and complex tasks a teacher has to face" [4]. In order to face these challenges, many design critique forms have emerged. Here, we briefly outline some challenges of both expert and group-based critique. Schon has written extensively about expert and learner interactions in the design studio, exploring the potential role of the expert. For example, although experts can provide useful insight, novice designers may feel evaluation apprehension and avoid sharing their work with experts [6]. Bennett has reported problems in giving individual feedback in large studio groups. In groups, critique can be a source of alarming humiliation for those whose outcomes have not yet matured [7]. Moreover, designers express dissatisfaction regarding the low quality of the average critique in the group context, despite the higher quantity received [8].

Thus, we notice that under each critique form, opposing forces can influence the success of a critique. In this paper, we wish to deepen the literature's understanding of creative critique by analyzing the roles of the *expert* and *group* more closely, and their effects on the general outcome of a student project.

**RESEARCH QUESTION**

We wish to investigate more deeply how varying critique form, with different *expert* and *group* conditions, affects higher-level critique goals. From the literature, we identify three central higher-level goals of a critique: (1) addressing a designer's doubts by answering their questions, (2) delivering concrete suggestions for a design and (3) boosting a designer's confidence. These three goals appropriately describe the typical way a critique should evolve. Now, consider the following hypotheses:

*Hypothesis 1 (H1):* Students will have more of their questions answered in a one-on-one with expert critique form. This hypothesis follows directly and intuitively from the critique form. In a one-on-one structure, there is no emphasis on getting the opinions of several critique members and so students are able to quickly field their pre-meditated questions. Thus the student has more control over the critique experience.

*Hypothesis 2 (H2)*: Students will feel that a one-on-one with expert critique form will offer the most concrete and applicable improvements. We motivate this intuition in two ways. Firstly, experts have been exposed to a large number of examples of the problems and solutions that occur in their domain and so are able to "access information in larger cognitive chunks than novices can, and to recognize underlying principles" [9]. In this way, an expert can direct a critique more effectively. Secondly, idea exchange in groups is an inefficient process resulting in productivity loss [10]. Thus, a one-on-one critique form will stimulate idea generation.

*Hypothesis 3 (H3)*: Students will feel most confident about their project following critique in the group with expert condition. We deem confidence to be dependent on both satisfaction with the current progress, and with the future direction of the student's project. At first glance, this hypothesis may seem counterintuitive given H1 and H2. However, we believe that validation by several critique members will spur the student on. Moreover, the presence of the expert will also structure the critique to yield maximal value from the session and thus boost a student's confidence with regard to future progress.

**METHOD**

To address our research goals, we studied student designers and experts who were participating in a class at Stanford University titled *CS 91SI: Digital Canvas: Intro to Visual Web Design.* The class meets weekly for 2 hours; the first half of each class is a lecture from the instructors, while the second half focuses on critiquing each of the students' works. The students were assigned a 4-week design project during the study's duration, in which they created a "skill-print" or interactive self-exploration website. Weekly milestones for this project ranged from thumbnail sketches to prototypes. There were three critique sessions to observe, during which the students presented the work they completed for the previous milestone for 5 minutes each. Each critique was recorded for later analysis and students completed two qualitative surveys (before and after the critique session). The pre- and post-critique surveys were designed to identify hypothesis-relevant differences that occurred as a result of critique.

**Participants**

We ran a total of 18 participants, 15 of whom were student designers and 3 of whom were experts from the course staff. Student designers had limited experience with visual design, while experts had extensive experience in UI/UX design and design critique. All participants ranged from freshmen to coterminal graduate students of varying majors. They were primarily Computer Science majors, as web development skills were a prerequisite for the class. Participants were evenly split among gender, with 7 male students and 8 female students, and 2 male experts and 1 female expert. Participants were randomly assigned to critique groups that persisted throughout the length of the class; these groups also happened to be largely evenly split among gender.

**Experimental Conditions**

The students were split into three critique groups (consisting of 4-6 students and 1 expert). Each week, each group was assigned to a critique form condition: the *group with expert* condition (control), the *group without expert* condition, or the *one-on-one with expert* condition. In the first condition, the student designers presented their work to their critique group while the expert facilitated discussion. In the second condition, the student designers presented their work to the critique group, but the expert was absent. In the third form, the student designer presented their work privately to the expert.

The study was a within-subjects study. Initially, the groups were randomly assigned a condition, and were rotated between each of the conditions over the course of three weeks. Thus, each student and expert experienced all three forms of critique once by the end of the study. We chose to rotate the groups rather than examine long-term improvements to students' designs, as rotation reduced discrepancies between experts' critique styles and other students' responses to designs. As the study was performed on a class, and students do not always have perfect attendance, critique group sizes were not necessarily constant.

**Procedure**

Before each critique session, student participants were required to fill out a pre-critique survey. In the pre-critique survey, students recorded what they wished the make, what questions they had for critique regarding their design and their level of confidence with their design on a 7-point Likert scale. Students then assembled in their critique group, and were assigned a critique form condition. The students then presented their work to the group and expert, the group alone, or the expert alone depending on their assigned condition. The students all presented sequentially and in a random order, and the critique of their work lasted 5 minutes each. Finally, at the end of the entire session, students filled out a post-critique survey. In the post-critique survey, students recorded what improvements could be made, what questions remained unanswered and their revised confidence level on the same 7-point scale.

**RESULTS**

From our aforementioned experimental design, our aim was to have significant results emerge from both a quantitative analysis of (1) the pre- and post-critique surveys and (2) a classification of verbal statements made during critique. However, an initial analysis showed that the surveys did not yield any statistical significance. We believe that this is a result of an insufficiently large sample size of 15 students and an insufficient number of trials. As such, we relied instead on qualitative implications of survey results and a hybrid quantitative-qualitative analysis of the classification framework's results. We will now briefly describe how both the pre/post-critique surveys and the classification framework were used to drive our findings.

**Pre- and Post-Critique Surveys**

The symmetric design of the pre- and post-critique surveys allows us to investigate our hypotheses. These surveys indirectly investigate the critique forms by recording how the critique affected key measures. By asking students the questions they had before critique and questions that remained after critique, we wish to answer H1: how many questions were answered by critique. By asking students to record improvements, we wish to see which critique form yields the most concrete improvements, thus answering H2. Finally, recording confidence levels before and after critique allows us to observe how each critique form affects a student's' confidence.

**Statement Classification Framework**

Through a preliminary analysis of pilot data that we collected before the study, we designed a classification system for students' verbal statements (see Figure 1). Unlike the surveys, which indirectly allow us to observe the effects of critique, this classification system directly records the statement types made during critique. This framework was used as a supplementary tool to the surveys. The system is influenced by Buxton's categorization of user feedback [11], and has four classes of statements: *comments* (facts or personal opinions), *suggestions* (concretely stated improvements to the current design), *questions* and *responses*. Including *responses* allows the framework to recursively take into account the dialogue nature of a critique. We further discretized comments with polarity (positive, negative, or neutral), suggestions as superficial ("You should change the colors") or substantial ("You should add contrast to help people see the difference"), and questions as guidance-seeking ("How should I change this button?"), feedback-seeking ("What do you think of this button?"), or clarification-seeking ("What does that button do?"). Finally, we also tagged suggestions as novel, piggybacked, or restated.

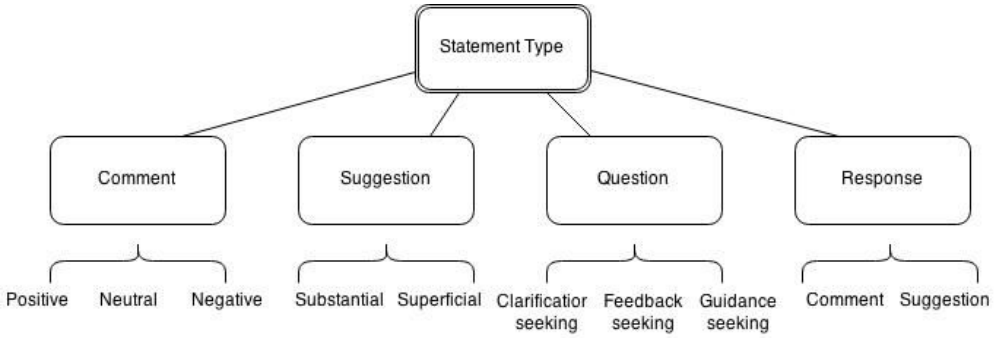


*Figure 1: A recursive statement classification framework*

Using this framework, two independent researchers classified the recorded statements. Classification disagreement was resolved with discussion. It is also important to note that we did not classify every sentence. In the course of conversation, some sentences were blended together, others were interrupted, and so on. Sentences with no content were skipped in our analysis, as were shorter utterances and exclamations (e.g. "Hmm", "Okay", and "Gotcha").

**Hypothesis 1**

To test our hypothesis, we planned to examine the difference in the number of questions pre- and post-critique, presumably measuring the number of questions answered by critique. However, we experienced problems in the cleanliness of our data: often times, students focused instead on minor implementation details that would not be answered easily through critique, and more often than not, critique often spawned more questions. Thus, it was not clearly evident that questions mentioned in

pre-critique responses were answered. It was inconclusive to say that students' questions were answered in critiques, let alone suggest that one-on-one critiques answered more questions for students versus group critiques.

**Hypothesis 2**
As with H1, we also had problems analyzing the pre- and post-critique responses to understand the concrete improvements that came about through critique. We planned to measure the difference in the number of improvements for any given student, conjecturing that any new improvements listed in the post-critique survey would have stemmed from novel suggestions during critique.

That being said, though it was not instantly clear from looking at the inputs and outputs of our critiques, we did a further examination of our classification framework by listening to recorded critiques and counting the number of suggestions raised during each 5-minute critique. Upon this examination, we listened to 17 different critiques of different forms (7 group with expert, 4 group without expert, 6 one-on-one), and counted the number of suggestions made. Performing one-way ANOVA, we found means for the number of suggestions raised: 5.7143 for group with expert, 4.5 for group without expert, and 3.1667 for group without expert. However, our p-value of 0.186 was still high due to our small sample sizes.

Statistics aside, we still think that there was an overall trend in that there were more novel suggestions in group critiques and fewer in one-on-one critiques. This is substantiated by observing the group with expert and group without expert conditions together to perform an unpaired t-test on the mean number of suggestions in group and one-on-one critiques. Performing the unpaired t-test yields means of 5.27 and 3.17 suggestions for group and one-on-one critiques, respectively. While the p-value of 0.0944 still did not meet the value of 0.05 for statistical significance, it is the most significant difference we've seen thus far in our research, and even more impressive when considering the small sample sizes.

**Hypothesis 3**
To test which forms of critique elicited higher confidence levels in students, we compared the mean change in confidence in the pre- and post-critique survey responses for students. In performing a one-way ANOVA on 20 completed responses (6 from group with expert, 4 from group without expert, and 8 from one-on-one), we noticed relatively similar means. The mean changes in confidence levels for the group with expert, group without expert, and one-on-one conditions were 0.625, -0.167, and 0.667, respectively. With a p-value of 0.496, our differences in means were not at all statistically significant, and no conclusion about different critique forms eliciting higher levels of confidence could be made.

**DISCUSSION**
There are a number of possible reasons why our results were not significant. Most importantly, we were only able to run three trials with a very limited number of participants, and could not control for differences among the three experts in their own personal critique styles. Despite the lack of statistical significance, we believe that the data we collected reveals trends in critique that we would not have otherwise considered. By exposing these trends, we are able to find new ways of understanding the factors at play in a critique session and enrich future studies on design learning and critique.

**Tradeoffs in Idea Generation**

We observed that more suggestions were presented in the group conditions as opposed to the one-on-one condition. However, the one-on-one condition seemed to yield a larger proportion of substantial suggestions. For example, a one-on-one session had only three suggestions, but each was well-developed and discussed by the expert and student designer in detail. In contrast, the group critiques often had rapidly presented suggestions that were not thought through, were silly, unreasonable or superficial. For example, in one group with expert critique session, the student designer presented a "sheet music" design. Suggestions for improvement included: "You should play music [when somebody clicks on a note]," "You should play one note," "You should play a song," "You can open a music player at the bottom," and "[You can play music] when the user hovers."

This is in line with previous research that suggests that group idea generation results in inefficiencies and social loafing [10]. We believe that there are tradeoffs to each kind of idea generation: students are able to either think through a solution deeply, or are exposed to a wide variety of ideas and perspectives in a brainstorm-like situation.

**Student Participation in Groups**

Between the group with expert condition and the group without expert condition, we noticed a trend that students participated more actively when the expert was not present. Very few novel contributions originated from students in the group with expert condition, while all novel contributions in the group without expert condition must have originated from the students. The students also noticed this trend in their own way. In the post critique survey, one student wrote that students "were more inclined to speak" when his group was in the group without expert condition. Another student noted that students "defer to the expert" in the group with expert condition. This suggests that when an expert is present, diffusion of responsibility occurs, and students feel less responsible for making substantial contributions. Additionally, because of the expert's experience, students may agree with the expert and not make their own judgments.

**Role of Expert's Critique Style**

Our data also suggests that the expert's own critique style may have heavily influenced the direction of conversation in the critiques. For example, one expert tended to provide more concrete directives than others did ("For next week, you should make a mockup of [...]"). Another expert asked students questions such as "What don't you like about this?", providing more room for negative comments. Features such as this may have affected students' expectations of critique and their roles within it. Because students always stayed in the same group with the expert, they may also have been influenced by their expert's style when they performed the critique themselves in the group without expert condition. While we believe that the experts' differing styles may have affected the outcome of our study, we also feel that it is an interesting feature to take note of.

**Limitations and Future Work**

Our data, and ultimately our results, were limited by factors such as time and the available setup for us to run our experiment on. We were only able to run three trials of our study over the three-week period; we feel that a longer study in which we can gather more data would yield statistically stronger results.

Additionally, certain aspects of the setup were unaccounted for in our study. For example, experts had varying communication styles that affected the direction of the critique, such as what elements of the critique they concentrated on (e.g. feasibility, aesthetic etc.). The students also may have started the critiques with varying levels of knowledge and skill that we did not account for.

Finally, our sentence classification framework, while thorough and based on pilot tests, prevented us from fully capturing the range of thoughts, ideas, and communications that occurred in a critique. Our transcription method omitted potentially meaningful vocalizations if they were short or ungrammatical utterances, such as a chorus of "yeah!"s or murmurs of confusion. Moreover, any transcription data is arguably ineffective in fully encoding the critique process, as it fails to recognize nonverbal communications such as body language, suggestions in the form of sketches or on-screen designs. In fact, Ökin and Lin [12] identify six different activities that occur in the design review and development process--thinking, examining, talking, writing and listening--and that people alternate between activities in rapid succession during novel design decisions.

Although our results yielded no statistical significance, we feel that designing a similar study that controls for the above factors and is run over a larger number of trials will address that issue. Further, our approach could be used to explore the features and outcomes of differing critique styles in a longitudinal study. With such a setup, we could perceivably gain insights on long-term effects of critique on improved student design skill, project improvement and performance over time and critique effectiveness.

## CONCLUSION

In this paper, we describe our efforts in identifying the role of the expert and the group in critique success. Through our exploratory study, we collected both quantitative and qualitative data, from which a number of interesting patterns emerged. While our quantitative data collected from student surveys before and after each critique session yielded no statistical significance, we were able to identify some general communication trends that occurred in each of the different critique styles. This included varying quality of idea generation in different critique forms, reduced student participation in the presence of an expert, and the influence of an expert's pre-existing critique style. These findings motivate future work in creative critique success and the roles of individual expert styles.

## ACKNOWLEDGEMENTS

We would like to thank the students and staff of CS 91SI for their willingness to take part in our study, and for their readiness to adapt to our changing plans. Finally, we would like to thank Michael Bernstein, Lydia Chilton, and the rest of the CS 376 staff for their continued support.

## REFERENCES

1. Taylor, M. J., & McCormack, C. (2006). Creative critique in visual communication: Effective feedback in design projects through voice-centred studio-based communication and online delivery. In *Thinking the future: Art, design and creativity. Proceedings from the 2006 australian council of university art and design schools (ACUADS) conference.* Retrieved from http://acuads.com.au/wp-content/uploads/2014/12/taylor_mccormack1.pdf
2. Feldman, E. B. (1994). *Practical art criticism*. New York: Prentice Hall.
3. Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53-64. doi:10.1080/03075070120099368
4. Bennett, R. (1997). Pushing Cows up Trees. In *Giving quality feedback to design students working within large groups. A paper to the Design and Education Council of Australia Conference*, 3-6 July 1997, Gold Coast, Australia.
5. Schön, D. A. (1983). *The reflective practitioner: How professionals think in action* (Vol. 5126). Basic books.

6. Marlow, J., & Dabbish, L. (2014). 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* (pp. 922-933). doi:10.1145/2531602.2531651
7. Horton, I. (2007). The relationship between creativity and the group crit in art and design education. *Creativity or Conformity*.
8. Luther, K., Pavel, A., Wu, W., Tolentino, J. L., Agrawala, M., Hartmann, B., & Dow, S. P. (2014, February). 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* (pp. 21-24). ACM.
9. Cross, N. (2004). Expertise in design: an overview. *Design studies*, 25(5), 427-441.
10. Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review*, 10(3), 186-213.
11. Tohidi, M., Buxton, W., Baecker, R., & Sellen, A. (2006, April). Getting the right design and the design right. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1243-1252). ACM.
12. Akin,Ö., & Lin, C. (1995). Design protocol data and novel design decisions. *Design Studies, 16(2)*, 211-236.