

Emoji as a medium for non-expert critique

Qing He
qinghe4@illinois.edu

Cassandra Jacobs
cljacob2@illinois.edu

Youshan Zhuang
yzhuang3@illinois.edu

ABSTRACT

In this paper we attempted to see whether emoji, a way of expressing emotions or sentiments in text messages, can augment or replace text-based critique provided by novices. Emoji may be a smaller vocabulary that is easier to acquire than more complex aspects of designs that may be difficult for novices to verbalize. 180 workers from Amazon Mechanical Turk provided critiques of three designs using emoji and text. We then assessed the independent contribution of emoji and text to judgments of quality (specificity and implementability) as well as the emotional reactions of the designers. We also gathered judgments from the designers about whether the emoji increased the quality of the text critique as well as whether the text increased the quality of the emoji critique. The results show that using emoji for critique is less useful than using text for critique, and that emoji used with text are not better than text critique alone, even among novices. However, the quality of emoji critique increased with the quality of the text, suggesting that emoji can be used to identify the same problems as text comments. We therefore conclude that crowdsourced critiques containing emoji do not necessarily improve the quality or interpretability of text comments, but discuss the potential uses of emoji in other contexts.

Author Keywords

emoji; critique; non-experts; design; crowdsourcing

INTRODUCTION

Crowdsourced critique can be difficult to elicit when the critiquers are not designers. Generally, novice critiquers in a crowdsourced system require considerable scaffolding to provide useful critique to designers, which in many cases may be expensive or time costly [?]. On the whole, design critiques take the form of verbal commentary, such as "The contrast between the text and the background should be higher." This style of verbal commentary may be difficult for non-experts to verbalize, even though non-experts identify many of the same problems in designs that experts do. Novices do have access to their own gut reactions to and initial impressions of a design, which can be harnessed to make the critique process easier. Cognitive style theories suggest that expressing moods and emotional reactions is important in eliciting good feedback [?]. If it is possible to provide feedback as well with emotional

reactions instead of with text, which requires a specific vocabulary, then images or abstract emotional content will lead to good, but efficient critique. Not leaving verbal commentary may reduce some of the performative aspects of critique [?] while having the same effectiveness as verbal communication [?]. Some have already done research along these lines, such as the MoodSource system of Robb et al. (2015), which used abstract images to assess the moods of participants [?].

Another alternative, however, is to use a system of emotional and conceptual expression that is already popular. Emoji, a type of Unicode character, are similar to emoticons and are commonly used in both American and international text message systems like Google Hangouts, WeChat, and iOS. Emoji are a growing part of modern culture, as seen by the White House using emoji in their use of Twitter and novels like *Moby Dick* being translated entirely into emoji ([?], <http://www.emojidick.com>). While emoji have been claimed to be able to express things that words do not, the use of emoji has been studied in intimate social contexts or on social media such as Facebook, Twitter, and Instagram. Emoji are restricted to certain social situations, often between close friends [?], but provide a conventionalized vocabulary of facial expressions as well as other symbols.

Instead of using a complex set of abstract images like those of MoodSource, we designed a critique system that employed emoji that would give novice critiquers a common and purely visual language that can cut across cultural boundaries and design environments. A potential extension of the emoji vocabulary to allow for non-experts to feel more comfortable about conveying their impressions of a work in progress could be useful and allow for non-experts to provide feedback without necessarily feeling like they are threatening the designer. To that end, we integrated the emoji vocabulary a simple critique system that allows users to place comments on specific locations on a screen as well as add emoji freeform to any element of a design.

EXPERIMENT DESIGN

Web emoji critique system

Emoji palette

As a first pass, we posted a task on Amazon Mechanical Turk that asked 40 workers to critique several sections of a powerpoint using 12 emoji that we felt conveyed emotions effectively along with a handful of more abstract ideas. We had them also provide text and emoji comments for three pre-defined regions on a simple powerpoint design. Then, the authors rated the relevance of the emoji to the text comments provided on a scale of 1-7. We then averaged the relevance scores, with many expressing the same ideas as the text comments. Some emoji were problematic because they did not match the commentary

provided. The seedling emoji, for example, was intended to suggest "this looks promising", while the magnifying glass is intended to mean "make this larger" or "look at this region". One emoji was not used at all.

Of this initial set of 12 emoji, we chose a small palette of eight emoji ranging from a very positive reaction (a face with hearts for eyes) to a negative reaction (a face that looks like it is in pain) plus three additional symbols (a bullseye, a yield or caution sign, and a question mark), since these emoji were most often related to text comments in design critique. It was also important to simplify the palette down to accommodate a response set that could be managed easily by novices.

Alpha emoji-text critique system

We chose firebase as our website host because it supports real-time updates in the cloud. Firebase stores data in JSON form, and use NoSQL data structure so that it is efficient in storing, retrieving, updating, and is highly scalable. Unfortunately, firebase does not support image storage, so it is hard to capture and store the snapshot of the emoji feedback. The final version of the critique system stores the position of each emoji relative to the design and generate new emoji images on the webpage to present the emoji feedback for designers. Designers have the option to view only the emoji, only the text, both, or neither when looking at design critiques.

The website is available at <https://design-feedback-emoji.firebaseio.com/>. Currently there are three designs with two versions, as well as the feedback that designers saw. In one version of the website, emoji must be placed onto the designs. In the other, emoji do not need to be placed onto the design. We believed this functionality would be useful for crowdsourcing critique with our system because in prior work, emoji were often not used or workers expressed that they would not use emoji otherwise. We discuss later the fact that with this interface, workers almost always used emoji in their critiques, suggesting that all that is needed is a sufficiently clear interface to elicit emoji-style critique.

Predicted results

Because emoji elicit a number of different possible emotional reactions, and because emoji are less specific than text, we predicted that emoji critique would not be very useful for designers on their own. We also predicted that the value of text would not change regardless of whether emoji were there or not, but that when emoji were considered with the text at the same time, the value of emoji would be additive with the value of the text. In other words, text augment emoji critiques, but emoji do not augment text critiques, at least in part because emoji without text critiques are difficult to interpret. We present our predicted results in Figure 1.

Procedure

Scoring and analysis

To evaluate the quality of emoji critique, the authors separately two designs: their own and a design of one of the other authors. We also evaluated the quality of text critique, as well as the impact both types of critique have on each other.

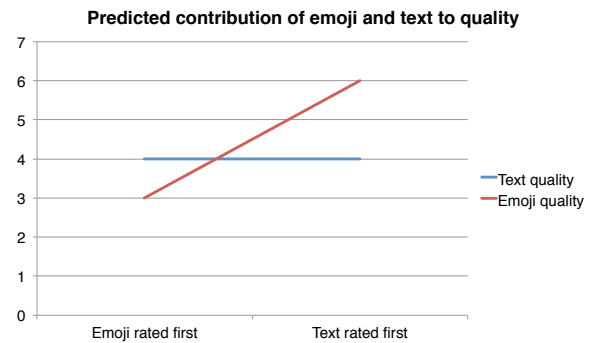


Figure 1. Predicted relationship between quality ratings of emoji and quality ratings of text. When text comments are used to interpret emoji, the quality of the emoji critique should be very high. When emoji critique are used to interpret the text, the quality of the text should not change from the text alone.

Quality scoring

The quality of critiques was rated on a scale from 1 to 7. We considered a number of different aspects of the critique, particularly the specificity of the comments (whether the comment identified a particular element of the design that could be changed or that was well executed) as well as the implementability of their suggestions for improvement. High-quality comments are therefore both specific and implementable. Here we provide an example of a high-quality comment from our experiments as well as a comment the designers identified as low quality. In response to the poster design, one critiquer said, "The green lettering in "Art & Design" does not seem to fit harmoniously with the "glitched" effect and purple/orange hues of the rest of the lettering." Similarly, in response to the resume design, one good comment read, "I do not like the name on two lines, nor the font color."

Quality rating procedure

Every design was rated by two designers – the original designer as well as one of the other authors of the paper. For half of the critiques, one designer would rate the quality of the emoji, while the other designer would rate the quality of the text. When they finished rating that half, they then evaluated the value of the emoji-text combination, either assessing the quality of the emoji in the context of the text, or the text in the context of the emoji. After the combined emoji-text rating, the two designers traded rating procedures. Both of these measures were taken to minimize the contribution of individual differences of the designers and their designs to the quality ratings, particularly as the authors enjoy emoji. The rating procedure is presented in the flowchart available in Figure 2.

Removal of low-quality feedback

Feedback were removed from the analysis when they did not contain one emoji that was placed on the design or when they did not contain text comments. Retaining the values would have made it difficult to assess the added value of emoji in conjunction with text. There were a resulting 31, 55, and 54 critiques for the app, poster, and resume conditions respectively. Of all of the design critiques we received, only eleven did not include any emoji: 4 for the app design, 3 for the poster

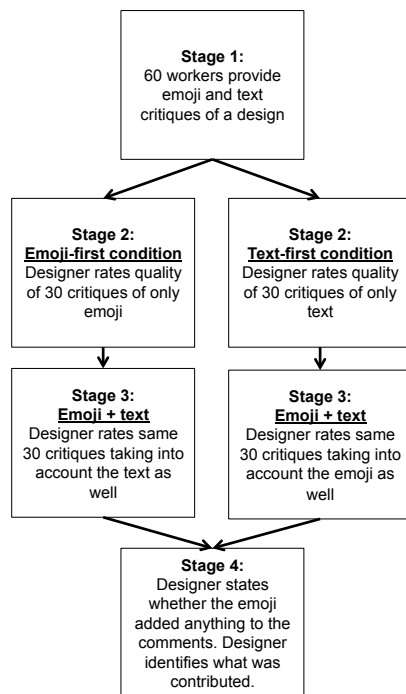


Figure 2. Order of ratings for each critique for a given design. Each design had 60 critiques. Half of them were rated only by their text first, while the other half were rated only by their emoji first, text comments were evaluated afterward for their quality in the context of the emoji. When text was rated first, emoji were evaluated for quality in the context of the text. This allows for us to identify the independent contributions of each medium of critique while also assessing whether emoji or text increased the value of the other medium.

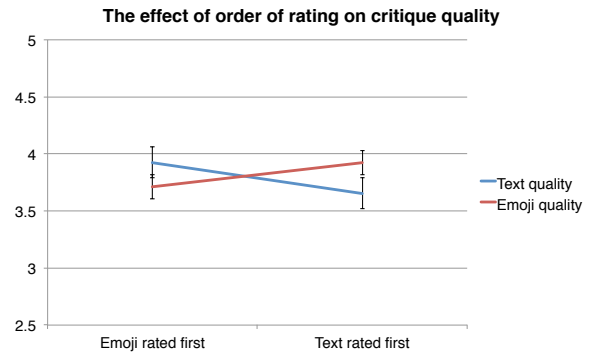


Figure 3. The quality of the four rating types for the app design. Emoji decrease the quality of text, but text increases the quality of emoji. Text ratings are higher than emoji ratings when they are presented alone.

design, and 4 for the resume design. That means, out of 30 feedback that emoji are not required, an average of less than 13% of people were not motivated to provide emoji. Most people are still willing to provide emoji feedback, and this may be because choose an emoji from a small palette and dragging it does not take a lot of effort.

Results

Quality ratings and the timing of information presented to designer

We constructed a linear regression model examining quality score as a function of which of the four conditions quality was assessed in. Each critique was rated for the value of its alone, the value of its text alone, the value of the emoji after processing the text, and the value of the text after processing the emoji. We predicted that emoji would significantly improve the value of text critique but not vice versa. The linear model compares all of the classes to our baseline condition, which is the text alone condition. Because this is the typical format of critique, it is important to see whether emoji alone provide less information than text. We found that text scores were generally very positive across all three designs (mean = 3.78), but that emoji alone were significantly worse critiques than text alone (beta = -0.53, $t = -3.08$, $p < .01$). Additionally, text did not significantly improve emoji critiques (beta = 0.23, $t = 1.33$, $p = .19$). However, emoji marginally worsened the quality of text critiques (beta = -0.33, $t = -1.91$, $p = .06$). Below we plot the results for the app, poster, and resume designs respectively in Figures 3, 4, and 6. We also discuss the different pattern of results we see for each of the different designs.

One particularly high quality emoji-based critique we provide here in Figure 5. Although the critique is short, it is very specific and gives clear instructions on how the designer should improve the design. The emoji clearly reiterate the workers' opinions: (1) Unhappy faces on the green and brown to show that the worker does not like the green and brown colors. (2) Smiley faces to show that the worker likes the font: A, &, D, I, and G. (3) An unhappy face is placed on 2014 to show that the worker does not like the box. Accordingly, the emoji feedback is likely to be useful as well and adds value to the text critique.

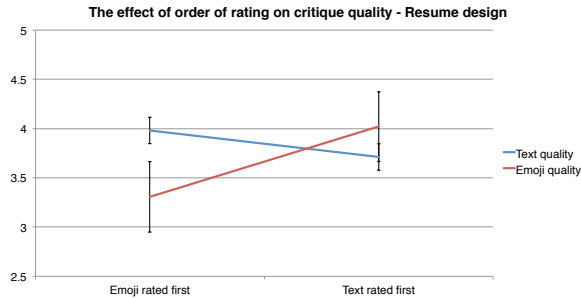


Figure 4. The quality of the four rating types for the resume design. Emoji are more effective when evaluated with text, but text is not more effective when evaluated with emoji. Text comments out of context are high in quality, but emoji comments are not, perhaps because the emoji are difficult to interpret out of context.



Figure 5. In this example, the worker provided high quality emoji and text critiques. Their text comment was specifically, "I do not like the green and brown colors. Love the font on the A&DIG. 2014 should not be in the box like that."

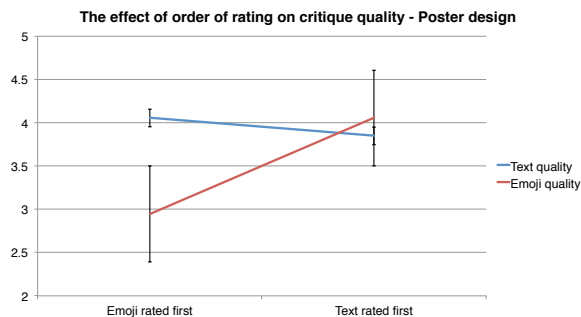
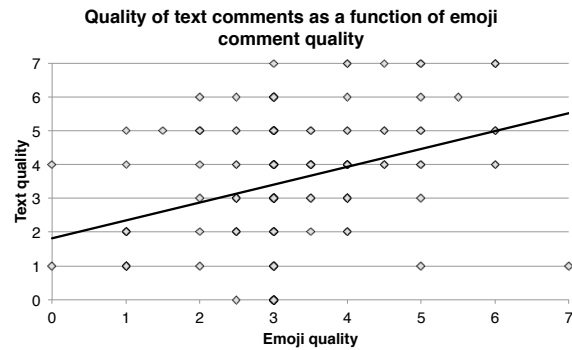


Figure 6. The quality of the four rating types for the poster design. Emoji are more effective when evaluated with text, but text is not more effective when evaluated with emoji. Text comments out of context are high in quality, but emoji comments are not, perhaps because the emoji are difficult to interpret out of context.

Figure 7. As emoji quality increases, so does the text comment quality. This suggests that good text comments will also come along with interpretable emoji comments.



Relationship between emoji and text quality ratings

We rated emoji and text separately from each other. One possibility is that the combined scores reflect how difficult the emoji are to interpret after having read the text and vice versa. Perhaps each domain critiques information well but the two domains are distracting when evaluated together. To test this possibility, we tested for whether there was a correlation between the quality of the emoji critique and the quality of the text critique. There was a strong and significant relationship between the quality of the emoji critique and the quality of the text critique, suggesting that good critique can come in both modalities (Pearson's $r = 0.40$, $t = 4.91$, $p < .001$). We plot this relationship below in Figure 7.

The result here is promising because it suggests that eliciting thoughtful text feedback might lead to better emoji feedback as well. This result is against our initial prediction that emoji would be easily used by novices. While almost everyone was able to use emoji for critique, good critique appears to be domain-general. That is, it would be faster to obtain emoji critique, but getting good emoji critique might still require teaching novices what details of a design are worth critiquing and how to provide such critique.

Discussion

We predicted that text critiques would make emoji highly interpretable and would increase the quality of emoji critiques. Instead, we found that emoji were not significantly more interpretable when they were interpreted in the context of text critiques, while we found that text comment quality actually decreased when the emoji were considered. However, there could be other reasons that lead to this result of low-quality on emoji comments. Many emojis were applied to random locations on the designs and designers did not understand the meaning behind the ones located in random places, made these feedback low-quality to designers. We find that workers from Amazon Mechanical Turk are diverse and they do not have the same understanding of the meanings of the emoji as designers may, based on our previous pilot. On this platform, workers write comments in order to get paid. So we consider that crowdsourcing might not be a perfect platform to use emoji comments.

At the same time, we see that high-quality text critiques are

also associated with high-quality emoji critiques. This suggests that an emoji critique system could continue to be developed that would lead to even higher quality critiques which would have a consistently higher quality like we observed for text. We discuss additional applications worth exploring for emoji critique, as well as other ways that emoji critique could be evaluated that might not have been considered in our initial analysis.

FUTURE DIRECTIONS

Improvements to the critique system

The current feedback system is an alpha prototype that only supports only the most basic functions without careful interface design. (1) It should alert workers when they drag the emoji out of the design boundary. (2) It should show a clear match between emoji and text allowing workers to add a hideable text box for each emoji, constructing a emoji-text graph, or tagging each suggestion point and linking every emoji to a tag. (3) The website should remember IP addresses so that a refreshed page is not necessarily treated like a new page unless the worker wants to start over. (4) It should have a basic quality filter such as word count. However, as we can see from the above examples, if the instructions are specific and clear, even if the word count is small, it is still possible to get high quality responses. (5) It should reject a worker's feedback if no text is submitted. (6) The design should be able to be enlarged and return to the default size. The emoji should be able to change its size to show the importance of the critique and also allow emoji to fit into smaller spaces to more clearly show where the emoji is intended to go. (7) A better user interface design is needed to encourage the workers to be more serious in providing feedback, because this makes the HITs more appealing to complete. (8) From the designer's perspective, the feedback summary page should filter results to show only high quality feedback, and should be capable of generating simple summary statistics of emoji use and comment properties like word count.

Friend, peer, or expert-level critique

We find that crowdsourcing might not be a perfect platform to use emoji comments. However, emoji comments might be appropriate in some scenarios, such as when critiquers and designers interpret emoji in the same way and when critiquers are patient enough to apply the appropriate emoji to the design. An emoji-based critique system could be effective among peers or friends who share similar backgrounds. Alternatively, emoji might work well in peer critique in the classroom for similar reasons.

Subjective ratings

We consider that the quality of a critique does not necessarily mean the critique is useful, especially for emoji comments. Apparently text comments, but not emoji, can express professional terms such as "kerning" in design, which improves the quality of comments. The usefulness of emoji comments can also be assessed by measuring the amount of time it takes designers to understand the critiques. During the rating process, we rated the quality of the comments as designers; emoji comments took less time to rate. For example, one worker wrote

several sentences to express a single idea that their impression of the work was positive, but a single positive emoji would have sufficed.

An additional component of usefulness is the subjective reaction designers have to the comments. In a future study, we plan to rate the subjective reaction designers have to emoji-only and text-only comments to see if emoji have a stronger influence on the designers' reactions than text comments. These subjective reactions could be rated in a scale of 1 to 7 point from negative to positive reactions.

Heatmap visualization

One interesting analysis of these data would be to consider the valence or emoji types and their particular locations on a design. While crowdsourcing might not be an ideal medium for the use of emoji, emoji may still provide information in aggregate by identifying problematic and successful aspects of a design. If, for example, many grimacing emoji across many users were placed in the upper left hand corner, we could translate that into a dark orange color to indicate that many novice critiquers thought that area could use improvement. Heat maps of emoji valence might also be more interpretable and easier to generate by using emoji than text, which would need to be parsed for their sentiment and identify the regions on the design that the comments are relevant to.

CONCLUSION

Altogether, our results show that emoji might be used to complement text critiques despite the fact that emoji do not appear to be all that useful on their own. The fact that emoji critique increased in quality when the corresponding text critiques were more useful suggests that the use emoji and text reflect similar information about a design when placed into the right hands.

REFERENCES

1. Daantje Derks, Agneta H Fischer, and Arjan ER Bos. 2008. The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior* 24, 3 (2008), 766–785.
2. Eisuke Ito and Takafumi Fujimoto. 2013. A proposal of intuitive and immediate emoticons system to do non-verbal communication with smartphones. In *Intelligent Systems Modelling & Simulation (ISMS), 2013 4th International Conference on*. IEEE, 335–339.
3. Ryan Kelly and Leon Watts. 2015. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design* (2015).
4. 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.

5. Herman Melville. *Moby Dick*.
6. David A Robb, Stefano Padilla, Britta Kalkreuter, and Mike J Chantler. 2015. Moodsource: Enabling Perceptual and Emotional Feedback from Crowds. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. ACM, 21–24.