Do Numeric Ratings Impact Peer Reviewers?

Catherine M. Hicks
The Design Lab
UC San Diego
cmhicks@ucsd.edu

Scott Klemmer
The Design Lab
UC San Diego
srk@ucsd.edu

C. Ailie Fraser
The Design Lab
UC San Diego
cafraser@ucsd.edu

Purvi Desai
The Design Lab
UC San Diego
pdesai@ucsd.edu

Abstract
Research suggests that online peer review can provide critical help to learners who would otherwise not be given individualized feedback on their work. However, little is known about how different characteristics of review systems impact *reviewers*. This extended abstract presents preliminary results from an online experiment examining how explicit numeric ratings change peer reviews. A between-subject experiment found that peer reviewers who were asked to generate a numeric rating as well as general feedback gave significantly more explanations and made more positive comments compared with reviewers who were asked to give general feedback only. These exploratory findings suggest the need to further examine how online peer review systems' affordances can impact the reviews given in these systems.

L@S 2015, Mar 14-18, 2015, Vancouver, BC, Canada
ACM 978-1-4503-3411-2/15/03.
http://dx.doi.org/10.1145/2724660.2728693

Author Keywords
peer review; peer assessment; online learning; writing

ACM Classification Keywords
K.3.1: Computer Uses in Education: Collaborative Learning.

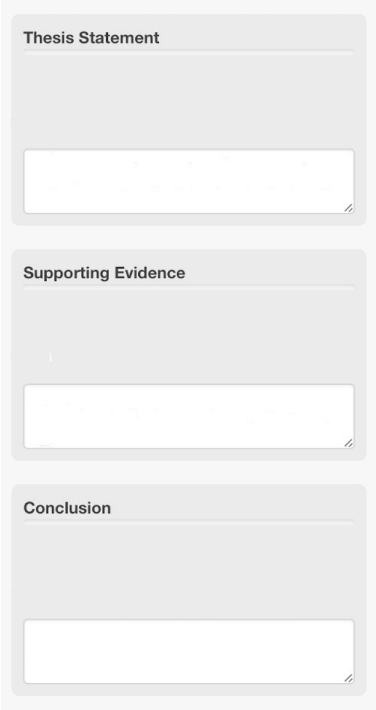

Figure 1. Non-numeric feedback condition. Peer reviewers were provided with open-ended text boxes only when giving feedback.

Introduction

Online environments provide increasing opportunities for people to interact collaboratively, creating and refining their work with feedback from peers. Different types of feedback may be formally imposed within an online system (such as voting on answers in an online forum or user-generated scores in an online course), or informally generated by group members (such as discussion between students in an online forum).

In the classroom, peer feedback has been shown to help students improve their writing as effectively as feedback from instructors [1, 3, 5, 7, 8]. It can therefore be a valuable replacement for expert feedback when the latter is not available, as in many online learning environments [e.g., 3]. Despite this increasing focus on peer review, little is known about how feedback givers are influenced by the structure of the feedback *environment*. Research on peer reviewers has primarily focused on improving reviewers' performance relative to instructors or experts; for instance, providing instructor training sessions for peer reviewers [e.g. 4]. Some recent research also finds that rubric clarity, such as providing parallel sentence structure and unambiguous wording, can improve the quality of peer reviews in online classes [3]. However, many questions about peer reviewers remain to be explored. For instance, in the absence of an explicit rubric, how are peer reviewers influenced by the affordances of an online environment?

GradStudio Study

We present preliminary results from a pilot investigation into online peer review for writing outside of a course environment.

Previous research has found that students are impacted by critical feedback, often making fewer writing revisions in response to feedback that criticizes their writing during early stages of review [2]. However, it is unknown whether reviewers generate different feedback when prompted to give more or less critical reviews. For example, would cues like a numeric scale change the content of reviews? In this paper we provide some suggestive work exploring this question. It is possible that reviewers will give more developmental feedback in the absence of critical cues, focusing less on what is wrong than on what can be changed. On the other hand, it is possible that providing peer reviewers with an explicit rating system encourages them to justify their ratings with deeper and less vague reviews.

Method

Participants

Participants were recruited through from online advertisements as well as flyers on a Southern California University campus. There were 53 participants; 70% applying to a Science program, 21% Social Science, and 9% Arts & Humanities or Other. 41% of participants were enrolled in college, and 88% of participants were native English speakers.

Materials and Procedure

Participants enrolled for the project through an online website created by the authors, where they provided their area of application, college enrollment, and whether they were native English speakers. Native and non-native English speakers were assigned in equal numbers to the two conditions in order to control for any language confounds across conditions.

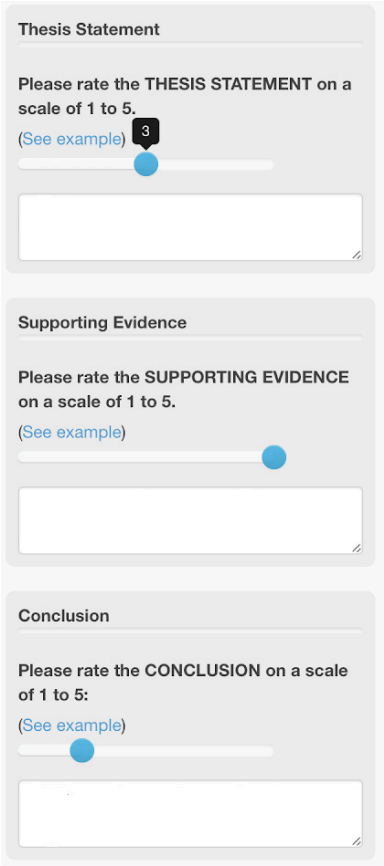

Figure 2. Numeric feedback condition. Peer reviewers were provided with numeric sliders and open-ended text boxes when giving feedback.

Participants then uploaded their graduate application essay for peer review through PeerStudio, an online peer review platform developed by Kulkarni and colleagues in the Stanford HCI group [6]. Participants were randomly assigned to one of two conditions: non-numeric or numeric. After submitting their essay, participants were required to review two other essays before receiving feedback on their own work. Based on this feedback, participants were then asked to submit their revised drafts for a second round of feedback. The study was open to enrollment and participation for one month, and participants were allowed to work at their own pace during that time.

Experimental Conditions

Participants were randomly assigned to receive and give reviews in one of two conditions: non-numeric (Figure 1) and numeric (Figure 2). Both conditions provided reviewers with open text boxes for comments on the essay's thesis statement, supporting evidence, and conclusion. In the numeric condition, reviewers were also asked to rate the thesis statement, supporting evidence and conclusion respectively with a sliding scale from 1 to 5.

Results

53 participants submitted at least one essay, and 24 went on to submit a revised essay after receiving feedback. A few participants submitted more than 2 drafts, with one submitting as many as 12.

Peer Reviews

204 reviews were submitted by peer reviewers. Because some reviewers failed to follow all the experimental instructions, the numbers of observations vary in the following analyses.

To analyze the content of the reviews, each review was assigned an "Explanation Score", representing the number of explanations given for suggested changes, and a "Positivity Score", representing the number of positive comments. One explanation point was given if a comment had a suggestion for improvement and justified it with an explicit explanation (e.g., "change the first paragraph *because it is repetitive and will be boring*"), while a half point was given if the review had a suggestion with an implied but unstated justification. No explanation point was given if a comment had no suggestions or made only vague statements (e.g., "change the conclusion"). Positivity scores were calculated by adding each positive statement in the review (e.g., "very strong intro", "fantastic writing"). Across 121 reviews, the mean Explanation score was 2.90, with a SD of 1.39, and the mean Positivity score was 1.54, with a SD of 1.31.

The presence of numeric ratings had a significant effect on the content of reviews; reviewers in the numeric condition were more likely to give explanations (M = 3.03) compared with reviewers in the non-numeric condition (M = 2.37), $F(1, 120) = 4.34$, $p = .03$. Reviewers in the numeric condition were also significantly more likely to make positive comments, $F(1, 120) = 4.55$, $p = .03$. However, Explanation scores and Positivity scores were not correlated.

Discussion

This study suggests that even small changes in the online review system, such as the presence or absence of numeric ratings, can influence the meaningful content of reviews. Peer reviewers who were given a numeric rating scale were both more positive and more detailed in their feedback, providing clearer

justifications for their suggestions and more encouraging comments.

One explanation for this finding could be that numeric ratings are perceived as more critical than open-ended comments alone, and peer reviewers therefore feel compelled to justify the implied criticism in their rating. It is also possible that requiring peer reviewers to choose a numeric rating encourages them to engage more deeply with the work in the first place, making more comparisons between essays. Finally, it is possible that providing an explicit rating system simply increased the overall clarity of the reviewers' task, which has been found to prompt higher quality reviews [3]. Future research should explore these possibilities.

Because of the relatively low number of participants and the self-selected nature of their participation, caution should be used in generalizing from these results. Follow-up work will investigate questions such as how receiving numeric ratings benefits or discourages online writers. We believe that our preliminary findings suggest useful directions for further research into how peer review systems' design impacts users both inside and outside of a classroom environment.

Acknowledgements

This work was supported by the National Science Foundation. We also thank Chinmay Kulkarni and the PeerStudio team for their support.

References

[1] Cho, K., and Schunn, C. D. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Comput. Educ. 48*, 3 (Apr. 2007), 409–426.

[2] Ferris, D. R. The influence of teacher commentary on student revision. *TESOL Quarterly 31*, 2 (July 1997), 315–339.

[3] 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 Trans. Comput.-Hum. Interact. 20*, 6 (Dec. 2013), 33:1– 33:31.

[4] Min, H. T. Training students to become successful peer reviewers. *System 33*, 2 (2005), 293-308.

[5] Patchan, M.M., Schunn, C.D., and Clark, R.J. Writing in natural sciences: Understanding the effects of different types of reviewers on the writing process. *Journal of Writing Research 2*, 3 (2011), 365–393.

[6] PeerStudio. http://www.peerstudio.org.

[7] Richer, D. L. The effects of two feedback systems on first year college students' writing proficiency. *College of Education University of Massachusetts-Lowell* (1992).

[8] Topping, K. Peer assessment between students in colleges and universities. *Review of Educational Research 68*, 3 (Oct. 1998), 249–276.