

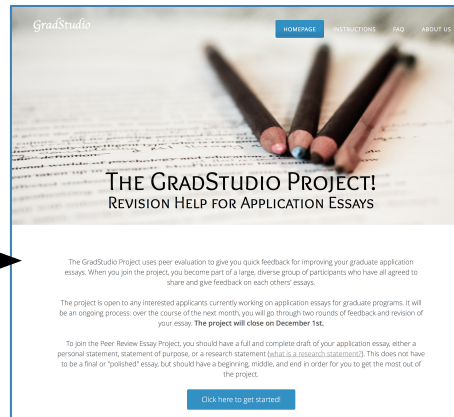
GRADSTUDIO: Exploring How Numeric Ratings Impact Peer Reviews

Catherine M. Hicks, C. Ailie Fraser,
Purvi Desai, Scott Klemmer

UC San Diego
The Design Lab

What is it?

- The GradStudio Project was an online experiment designed to use peer review techniques from online education in the context of graduate school applications.
- The goal: To examine how the presence or absence of numeric ratings affects the content of peer reviews.



Why do it?

- 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.
- We want to know: Do cues like a numeric scale change the content of reviews?

1

Participants signed up for the project through our website, where they completed a short background survey.

3

Participants uploaded their graduate application essay to PeerStudio for peer review.

2

Participants enrolled in our assignment on PeerStudio, an online peer review platform developed by the Stanford HCI group (www.peerstudio.org).

4

After submitting their essay, participants were required to review two other essays. Participants were randomly assigned to one of two conditions for this: **non-numeric** (left) and **numeric** (right).

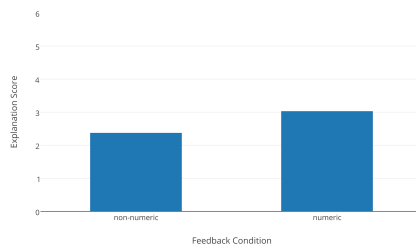
5

Once participants had reviewed two other essays, they could view their own feedback.

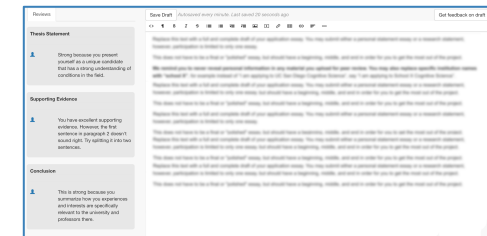
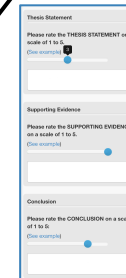
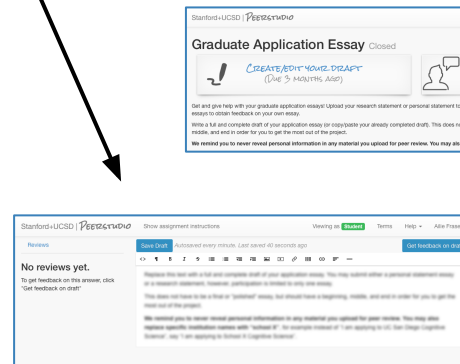
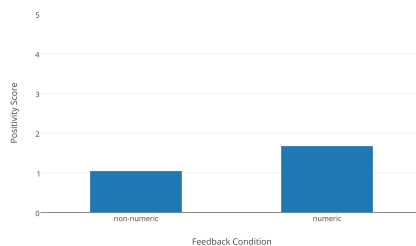
What happened?

53 participants submitted at least one essay for peer review.
204 reviews were submitted in total by peer reviewers.

Mean Explanation Scores of Peer Reviews by Feedback Condition



Mean Positivity Scores of Peer Reviews by Feedback Condition



What did we analyze?

- 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.
- Reviewers in the numeric condition were more likely to give explanations than reviewers in the non-numeric condition, $F(1, 120) = 4.34, p = .03$.
- Reviewers in the numeric condition were also significantly more likely to make positive comments, $F(1, 120) = 4.55, p = .03$.

What does it mean?

- 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.
- 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.
- It could also be that providing an explicit rating system simply increased the overall clarity of the reviewers' task.