# Indecent exposure: revealing peer reviewer's course performance can affect author perceptions of feedback

**Yasmine Kotturi**
Design Lab
UC San Diego
ykotturi@ucsd.edu

**Chen Shen**
Computer Science and Engineering
UC San Diego
c4shen@ucsd.edu

**Chinmay Kulkarni**
HCII
Carnegie Mellon
chinmay@andrew.cmu.edu

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

## ABSTRACT
Peer assessment has many strengths: it enables students to see their peers' work, share ideas, and give feedback. However, perceived peer assessment invalidity can weaken the ecosystem: students struggle to trust the grades that they receive from, and give to, their peers. In this paper, we explore shifting from the current status quo of random peer reviewer assignment to assignment based on student previous course performance. We then disclosed reviewers' course performance for 50% of authors. We find lower satisfaction of feedback and grades when the absolute value of author-reviewer course performance differences is nontrivial (a whole letter grade). In addition, authors who receive feedback from someone who is greatly out-performing them, and they know it, report the lowest satisfaction across all categories. We illustrate these finding in one in-person 20 student class on interaction design. Measures of efficacy include student perceived fairness of grades, perceived quality of feedback, and likeability of both grades and feedback, as well as actual quality of feedback.

### Keywords
Peer assessment, peer review, peer learning

## INTRODUCTION
Peer assessment can be pedagogically powerful: it exposes students to their peers' ideas, enables feedback and can enhance the quality of resulting work. Specifically, peer review can also catalyze fast, formative feedback on in-progress open ended work at a massive scale [5]. Such feedback and iteration are important for mastery learning [4]. However, one current challenge in the peer assessment ecosystem is that novices grade novices, whose biases can impact their grading ability. For example, students grade other students from their country on average 4% higher than students from a different country and peer grades are 7% higher than what grade an expert would assign [6]. While novices can often offer insightful feedback, the grading aspect causes this pain point.

Moreover, this pain point is two fold: not only can the feedback and grade giving lead to friction, the feedback and grade receiving can cause discomfort, too. Feedback from peers is not perceived well, as compared to feedback from figures of authority [3,7]. From preliminary data collection, we learned that students do not necessarily trust the grades that they receive from their peers: "Why should someone who knows just as much as me be grading my work?" And also "How can I be qualified to grade someone who I know just as much as?" In both directions, trust of oneself and one's peers is an issue.

There have been different approaches to tackling this challenge. For example, Calibrated Peer Review first trains students' review and grading techniques with controlled assignments [1]. After this calibration, they are able to grade their peers' work. This process, however, can be time intensive and can take away time spent learning the material. Another approach has been to try to scaffold the peer grading experience as much as possible through user interface design interventions. For example, Kulkarni et al created fortune cookies, helpful prompts to guide student comments, and also gave feedback to students on their grading biases, which improved peer grading accuracy [6]. Ordinal grading techniques, where students simply rank a subset of submissions, provides an alternative solution by removing the numeric grading required by novices altogether [9]. The above differing techniques have two aspects in common: 1. enhancing peer reviewer abilities by either making them better reviewers or making the system more natural for a novice mindset, and 2. random assignment of peer reviewers.

### Alternative approach to strengthening peer ecosystem
The above techniques, however, explore neither differences in how the feedback is framed, nor specific assignment of peers. There are take several strategies for optimal group formations: it is strongly recommended to ensure that students are exposed to those that are performing stronger [8]. Can we therefore apply this approach to peer assessment, leveraging the expertise of the stronger students, and help those who are struggling the most?

In addition, the framing of feedback is immensely powerful and can change how receivers view their feedback [2]. If we expose the previous course performance of the reviewers, does this affect how reviews are perceived? If so, does this effect differ if reviewers' performance is higher vs lower than author performance?

We hypothesize that disclosed course performance will lead to increased perceived fairness of grade, perceived quality of feedback, and overall satisfaction when the reviewer is outperforming the author (and opposite if under performing author).

## METHOD

### Design of Who-Reviews-Who Algorithm

Given a small population (n=20), we developed a high resolution algorithm to assign student reviewers. To begin, we first read in all students' current grades and sort in descending order. When assigning Reviewer #1, or the lower-performing reviewer, we duplicate the student list and shift the copied list one position up so that each student is receiving a grade from the person one position below, except the student with the lowest grade receives a review from a top student. When assigning Reviewer #2, or the out-performing reviewer, position down instead so that each student is receiving a grade from the person one position above, except the student with the highest grade receives a review from a low rank student. Due to the simplicity of the algorithm, Reviewer #1 and Reviewer #2 are reversed for the two students who received the best and the lowest grades on the previous assignment. By shifting the student list, we are guaranteed that each student will not grade the same person twice. For generating Reviewer #3, we created a set of possible partners from the student pool for the current individual A, randomly select one student B from the set and remove B from the student pool. However, if the student pool is empty, we then scan the list of pairs found and swap A from an individual C who is paired with someone not in A's history and still has possible partners. We do this repeatedly until no one pair can be found.

| Your student number | **Mastery (5pts)** | **Proficient (4pts-3pt** | **Weak (2pts-1pt)** |
|---|---|---|---|
| **Takes into account user goals** | Takes into account potential user goals and user questions are answered with clarify | Takes into accout user goals, majority of user questions are answered with information provided | Lacks the understanding of user goals/questions: neither are addressed, or done so weakly |
| **Provides actionable information** | Provides the user with an understanding of the big picture: what tasks can be accomplished by users via actionable information | Provides the user with an understanding of the big picture and what tasks can be accomplished by potential users, but not in an actionable manner | Does not provide the user with an understanding of the information, and by extension is not actionable |
| **Visual Style** | Highly visually appealing: consistent color scheme, layout and typography, has "WOW!" factor | Visually appealing: consistent color scheme and typography, but lacks "WOW" factor | Not visually appealing: lacking consistent color, layout, typography |

**Figure 1: Peer reviewers were given a brief rubric to help grade and elicit feedback: above is the dashboard rubric**

**Figure 2: An example of a student submission: an alert (left) and dashboard (right) user interface design**

### Implementation of Algorithm and Reflection Questions

We implemented our algorithm in one in-person 20 student class on Advanced Interaction. Each submission was assigned three reviewers. Students were given a rubric to aid their reviewing process (See figure 1). We did not provide any other type of grading assistance or scaffold, as we wanted to be as hands-off as possible during the review process in order to focus on the potential effects when framing of receiving feedback. Students provided feedback and grades for a dashboard and alert user interface design (Figure 2).

We intended to keep all stages anonymous. However, students presented their work in class before the reviews took place, and as it was a small class, most reviewers recognized their peers work. The reviews and grades, however, were anonymized.

Next, students returned the submissions they revised. We then resorted submissions based on author, and wrote on the the rubrics "Reviewer #1", "Reviewer #2", or "Reviewer #3": Reviewer #1s were reviewers who scored lower than authors on a previous course assignment, #2 higher or equivalent, and #3, random.

We exposed all three of the reviewers' previous course performance for 50% of the students: "This feedback is from a student who received a [reviewers score]/24 on the previous course assignment" was written on each rubric if in exposed condition. We assumed most students would remember what they had received on the pervious course assignment, so the relative performance (higher or lower) of the reviewer, compared to the author, was therefore implied.

Lastly, we asked students to complete a reflection form to elicit their perceptions of reviews (Figure 3). We asked each student to complete six elements per each reviewer, using a 5-point Likert scale. We ensured to differentiate perceptions towards grades versus feedback. We also encouraged general comments were encouraged about each reviewer, as well the overall peer review experience.


**Figure 3: A subset of the reflection questions we asked students to complete after they received their three peer reviews.**

## RESULTS

As mentioned, 20 students participated in the peer assessment stages: 16 of which participated in all steps of our intervention, therefore we report on data from these 16 students.

We categorized authors into two groups based on whether the reviewers course performance was exposed or not. Then we split reviewers into three categories: reviewers with higher relative performance, reviewers with lower relative performance, and reviewers of the same relative performance (tie in previous assignment grades). We then looked at the differences in self reported metrics across these two conditions and three sources. Interestingly, no stringent differences emerged from this categorization of data. Therefore, we parsed out instances where author-reviewer performance differences we nontrivial, a letter grade or more.

This highlighted two interesting findings: 1. There are negligible differences in author perceptions' when reviewers' course performance is equivalent, less than a letter grade higher, and less than a letter grade lower than the author, whether reviewer performance is exposed or not (Figure 4). However, 2. when reviewers' and authors' course differs by at least one letter grade, either higher or lower, we see that overall perceptions of quality and fairness of reviews are lower (Figure 4). Moreover, authors who receive feedback from someone who is greatly out-performing them, and they know it, report the lowest ratings all both categories.

In short, we find exposing reviewer performance leads to lower satisfaction (satisfaction defined as perceived fairness

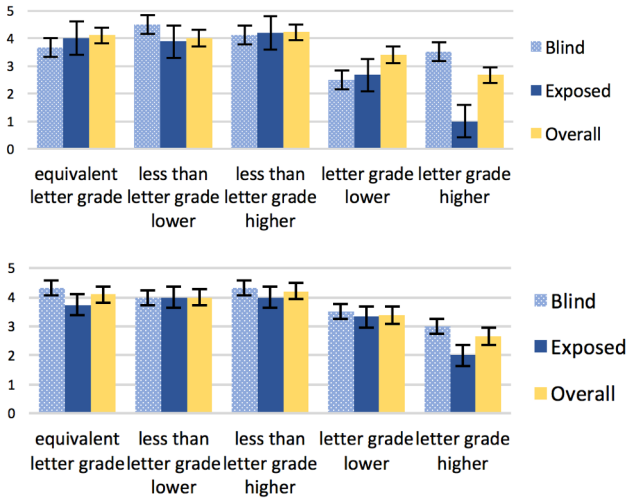


**Figure 4. Perceived quality of feedback (top) and perceived fairness of grade (bottom) is lower when reviewer-author performance differs by at least one letter grade. Y-axes represents Likert scale ratings, X-axes reviewer performance relative to author.**

and quality of review aggregate averaged) of feedback and grades when the absolute value of author-reviewer course performance differences is nontrivial (a whole letter grade). The likeability ratings echoed the fairness and quality ratings, thus gently reaffirming our definition of satisfaction.

When we compared quality of feedback with actual quality of feedback (which we manually coded on a 5-point scale, blind to condition), little to no consistencies emerged: actual quality of feedback did not correlate with perceived quality of feedback.

Instead, the seemingly largest factor in whether an author perceives feedback as high quality is, perhaps unsurprisingly, what grade they received: higher the grade, the more satisfied (Figure 5).

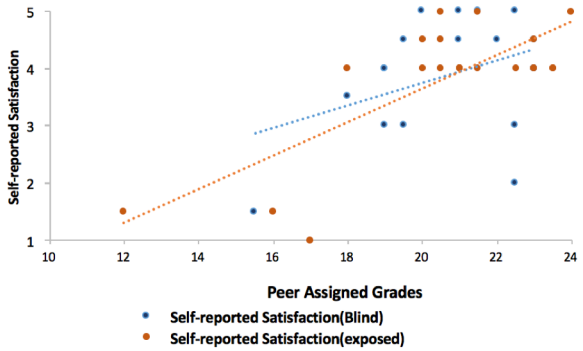


**Figure 5. Satisfaction (perceived fairness and quality of review aggregate averaged) is higher when assigned grades are higher, seemingly regardless to whether students know reviewer performance or not.**

## DISCUSSION

These results suggest that students may be sensitive to peer feedback from those who are strongly out-performing them. Therefore, if reviewers' previous course performance is exposed, it should be done with delicacy to prevent indecent exposure. For instance, we could instead write on rubrics "This feedback from a student who scored higher than you on the previous course assignment", and leave out altogether the exact grade.

### Future work

We see many opportunities to expand and build upon this work. For instance, we could implement a profession grader into the peer assessment process in order to offer expert level feedback to the top performing students.

Importantly, one weakness of our current study is that perceived quality of feedback is slightly higher perhaps when not potentially deserved: just because a reviewer has a higher grade doesn't necessarily mean that their feedback is actually of higher quality. Therefore, this increase should be merited in that the quality of feedback is increased: we believe introducing a professional grader in the peer assessment process will have an aggregated beneficial impact on students' grading abilities as the expert can offer more guidance and structure, and provide great examples of good reviews and grades.

In addition, there are other metrics that could be used instead of previous course performance to decide who is assigned to who: for instance, course participation. And lastly, it is important for us to explore with this algorithm in a larger class, in order to ensure effects are coherent.

## CONCLUSION

While peer assessment has many strengths, perceived invalidity of the system can weaken the ecosystem: students struggle to trust the grades that they receive from, and give to, their peers. In this paper, we explored shifting from the current status quo of random peer reviewer assignment to assignment based on student previous course performance. We also disclosed reviewers' course performance for 50% of students. We find lower satisfaction of feedback and grades when the absolute value of author-reviewer course performance differences is nontrivial (a whole letter grade). In addition, authors who receive feedback from someone who is greatly out-performing them, and they know it, report the lowest satisfaction across all categories. This therefore suggests that if reviewers' previous course performance is exposed, it should be done with delicacy, and students may be sensitive to peer feedback from those who a strongly out-performing them.


## ACKNOWLEDGMENTS

We are grateful to the students who participated and gave us insightful feedback on our peer review system. We also thank Ailie Fraser for feedback throughout the development of the algorithm and Chinmay Kulkarni for feedback on the reflection form.


## REFERENCES


1. Carlson, P.A. and Berry, F.C. Calibrated Peer Review and assessing learning outcomes. *Frontiers in Education Conference*, (2003).
2. Cohen, G., Steele, C., and Ross, L. The Mentor's Dilemma: Providing Critical Feedback Across the Racial Divide. *Personality and Social Psychology Bulletin 25*, (1999), 1302–1318.
3. Fedor, D., Davis, W., Maslyn, J., and Mathieson, K. Performance improvement efforts in response to negative feedback: the roles of source power and recipient self-esteem. *Journal of Management 27*, (2001), 79–97.
4. Guskey, T. Closing Achievement Gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery." *Journal of Advanced Academics 19*, 8-31 (2007).
5. Kulkarni, C., Bernstein, M., and Klemmer, S. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. *Learning at Scale*, (2015).
6. 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 Transactions on Computer-Human Interaction (TOCHI) 20*, 6 (2013), 33.
7. Leung, K., Su, S., and Morris, M. When is Criticism Not Constructive? The Roles of Fairness Perceptions and Dispositional Attributions in Employee Acceptance of Critical Supervisory Feedback. *Human Relations 54*, (2001), 1155–1187.
8. Oakley, B., Felder, R., Brent, R., and Imad, E. Journal of Student Centered Learning. *Journal of Student Centered Learning 2*, (2004).
9. Raman, K. and Joachims, T. Bayesian Ordinal Peer Grading. *Learning at Scale*, (2015).