Score-Group Framing Negatively Impacts Peer Evaluations

CELIA DURKIN, University of California, San Diego, USA
FEDERICO ROSSANO, University of California, San Diego, USA
SCOTT KLEMMER, University of California, San Diego, USA

How does group membership framing affect the feedback students provide learners? This paper presents two between-subjects experiments that investigate the effect of *Ingroup/Outgroup* membership on effort spent in peer evaluations, and whether the group membership criterion affects quality and stringency of evaluation. Two peer-review assignments were implemented in two separate classes. In the first study, students were nominally grouped by location they sat in class and non-nominally grouped by current class score; each was asked to review an *Ingroup* and *Outgroup* peer assignment. A second study randomly assigned students to one of four group types (random, score, motivation, and location); students reviewed two *Ingroup* assignments. In both studies, score-grouped students graded their peers more stringently than students grouped by location. These studies illustrate for system designers the impacts of group framing – and the disclosure of that—in peer review tasks.

KEYWORDS
Peer feedback; minimal group theory; Human factors; experimentation

ACM Reference format:
Celia Durkin, Federico Rossano, & Scott Klemmer. 2018. Score-Group Framing Negatively Impacts Peer Evaluations. In *Proceedings of the ACM on Human-Computer Interaction,* Vol. 2, CSCW, Article 45 (November 2018). ACM, New York, NY. 19 pages. https://doi.org/10.1145/3274314

1 INTRODUCTION

Many classes—especially online--implement peer review to make open-ended assignments practical when the number of students outstrips the available grading labor from the teaching staff [17]. Peer feedback offers many benefits to learners: from providing exposure to new problems to instigating social interaction, which may lead to new understanding. However, the quality of peer review can be uneven [7,16,18], given the wide range of learner backgrounds, knowledge bases, and experience giving feedback. Attempts to standardize high quality feedback and scaffold the feedback experience have focused on framing task goals [17]and speeding [19].

In an online learning environment, the lack of physical connection and potential for anonymous submission creates social distance between peers [18]. Quality of peer review thus suffers due to a lack of intrinsic motivation and social accountability [28].

Being accountable improves feedback quality. When peer reviewers know that they are being evaluated, the quality of peer review improves [23]. However, evaluating the reviewers involves further expenditure of resources. Revealing identities of reviewers can cause reviewers to be influenced by fear of reprisal or a desire to please which has both beneficial and detrimental outcomes; the quality of assessment is not necessarily better [39].

2573-0142/2018/November – ART45 $15.00
 https://doi.org/10.1145/3274314

Like any collaborative working environment, successful peer review benefits from social closeness. One must be motivated, either intrinsically or extrinsically, to spend time reviewing someone else's work, and also trust that they will get quality feedback on their own work [23]. Social accountability is a type of extrinsic motivation. We investigate techniques for decreasing perceived social distance that intrinsically motivate reviewers.

1.1 Decreasing Social Distance Online

Increasing awareness of another's presence, whether visually or semantically, can foster a feeling of community online. Manipulating onscreen visual presence [12], sharing information about features of interpersonal similarity [36], and presenting information about group activities and group-based competition [29] all increase the commitment to their respective online communities.

We investigate whether forming similarity-based *Ingroups* decreases social distance and increases quality. Lightweight groups are extremely easy to implement, making this an attractive option if it is effective. Based on prior work [37], we hypothesized that grouping by a feature of similarity would elicit an *Ingroup* effect of increased commitment to helping another student. What type of features of similarity might enhance this effect? Could groups be arbitrarily assigned and still elicit an *Ingroup* effect [34], or do groups need some meaning? Does an *Ingroup* effect still occur without a salient *Outgroup*?

1.2 Minimal Group Paradigm

Psychological research on social distance finds that people allocate more resources to others who they perceive to have something in common with themselves, even when the commonality is arbitrary [34]. In a seminal study on intergroup behavior, Tajfel and colleagues asked subjects to complete a computerized task, estimating the number of dots on a screen. The subjects were then artificially grouped into two performance types: "over-estimators" and "under-estimators." Subjects were given a list of participants who performed similarly on the same task (without any other information) as well as a list of those who performed differently on the task from the participant. When asked to allocate money to participants in both groups, subjects allocated significantly more resources to participants in the group that had performed similarly on the task. In a control condition, where participants were not told anything about performance, money was more equitably allocated. Tajfel and colleagues performed a similar manipulation of groups—this time with nominal divisions by taste. Subjects divided people into groups based on preference for the artist Wassily Kandinsky or the artist Paul Klee allocated more resources to those in their taste *Ingroup* [4].

Since Tajfel's pioneering work, group paradigms have continued to be used in Social Identity Theory, reiterating that people tend to favor their own group when given a salient "other" group [1]. Even trivial connections can create a sense of belongingness and increase motivation in group work [29,32,34]. People evaluate those in their group more positively, reward those in their group more generously, and work to accomplish *Ingroup* goals more diligently [11,31,32].

1.3 Motivating Peer Assessment with Minimal Groups

Since nominal group identity can increase resource allocation, might it increase the quality of peer reviews? We performed two studies that sought to decrease this social distance by implementing nominal and grouping strategies in student peer review assignments. Our first study asked: Do nominal groups, consequential groups, or both increase quality and effort in reciprocal social computing tasks like peer review? Our second study asked: Can we elicit an *Ingroup* effect without the presence of a salient *Outgroup*?

Adopting methods from minimal group theory, our first experiment grouped people together by different features—a non-meaningful feature (seating Location in the classroom) and a task-relevant feature (current course performance), and measured effort spent and quality of peer review. If minimal group theory applies to peer review, we would expect to see more effort allocated to the *Ingroup* in both conditions, even though the location condition grouping feature was meaningless. However, results from the first study found no *Ingroup*/*Outgroup* differences in the Location Condition, and actually biases *against Ingroup* reviews in the Score Condition. Students grouped by Score spent less effort on peer review compared to students grouped by Location, and graded their peers lower.

We then surmised that Location can be a factor in social closeness [3,24], and thus may not have been a nominal (meaningless) group. In fact, the *Outgroup* could have elicited feelings of social closeness by drawing attention to a shared physical location (even the *Outgroup* students still sat in the same classroom). In order to control for this, we performed a second study, which included a control group. In the second study, we investigated whether knowledge of Location increased effort or knowledge of Score decreased effort, and whether the reverse-*Ingroup* effect of Score remains without an *Outgroup*. Our results showed that grouping students by Score did not significantly affect effort spent on peer review when no *Outgroup* is present. However, Score grouping still negatively biased peer grades, even absent an *Outgroup*. Our results indicate that students become more competitive with those who have a similar score in the class, and thus grade those peers harsher and allocate less effort toward helping those students. Over two studies, in which we grouped by Location, Score, and used an extrinsically motivational script, Score was the only grouping that affected effort spent and grading on review.

2. EXPERIMENT 1: DO MINIMAL GROUPS MOTIVATE PEER REVIEW?

The first study applied *Ingroup*/*Outgroup* theory to peer review to find out if artificially created groups affect quality and time dedicated to peer review. In addition, we tested whether the group-defining feature mattered in creating an *Ingroup*/*Outgroup* effect. We tested two different groups—one based on score and one based on classroom seating location. Score grouping—inspired by Tajfel's original study--highlights mental similarities or dissimilarities; location grouping does not—and thus should not bias ideas about performance. We investigated whether students allocate more time and effort to reviewing *Ingroup* submissions—in both the Location and Score conditions.

Knowledge of a peer's score affects both the reviewer and the author. Perception of an author's score can negatively affect a peer's motivation to do well on their own tasks—as reviewing peers with higher levels of performance can discourage students. In a study where students in an online course were asked to review peers' work, a large proportion of those exposed to exemplary examples quit the course [29]. Given that perception of score may be a powerful factor in perception of peers, we chose it as a non-nominal group, and divided students using a median split of their current score in the class. We hypothesized that students would allocate more resources to the reviews from their same score group.

The second grouping type was based on seating location in the classroom. Seating location in the classroom was chosen as a nominal(minimal) group, as seating location should not bias perception of peer's intellectual work. *Ingroup* bias in the Location Condition would thus suggest that Minimal Group Theory (grouping based on arbitrary differences, such as color) can be applied to peer review.

We hypothesized that students would allocate more resources to the *Ingroup* in both conditions, and the *Ingroup* effect would be stronger for the meaningful group—the Score Condition.

2.1 Hypotheses

1. A main effect of Review Type, such that students would spend more effort on *Ingroup* than *Outgroup* submissions.
2. No main effect of Condition: Students from Location Condition and Score Condition will spend the same amount of effort giving feedback on submissions.
3. An interaction between Condition and Review Type, such that difference between *Ingroup* and *Outgroup* will be smaller in the Location Condition than in the Score Condition.

3 METHODS

3.1 Participants

Participants were 101 college students (ages 18-22) enrolled in an "Introduction to Psychological Methods" course at a western US research university. All subjects were recruited voluntarily by class announcement. Recruitment and testing of subjects was approved by the university's institutional review board.

3.2 Procedure

Subjects were told in class (four weeks prior to the due date of their final assignment) that they could volunteer to participate in a peer-review exercise—where they would review two peers' drafts of their final paper, and in exchange have two peers review a draft of their final paper. The final paper consisted of a five-page proposal for an experiment, comprising Introduction, Methods, Results, and Discussion sections. Students were randomly assigned to one of two conditions: Score (N = 49) or Location (N = 52). Upon turning in their drafts, the students in the Location Condition were asked to self-identify which side of the class they usually sit on (left or right), and the students in the Score Condition were asked to provide their current score in the class, which could be easily looked up on the course website. They then reviewed two submissions before they received feedback on their own submission.

3.2.1 Submission and Distribution of Drafts. Students submitted their paper draft via Google Forms, and in return were emailed links to an *Ingroup* and an *Outgroup* submission, as well as a link to a rubric for evaluating them. The only information reviewers were given about peer submissions was their participant number and whether the submission was In- or Out-group. The order of In- and Out-group submissions was counterbalanced.

Students in the Location Condition reviewed an *Ingroup* submission from a peer who usually sits on the same side of the class, and an *Outgroup* submission from a peer on the other side. The email labeled each link as "a submission from someone who sits on the same (or other) side of the room as you."

Name:
Does this author sit on the same side of the room as you? Yes No
Order of Review (is this the first or second submission you have reviewed?)______
Rubric Instructions:
Please answer the questions (highlight or underline yes or no) and give comments/suggestions below where appropriate. You may write as much in the comments section as you want.

Intro Section:

Is the research question falsifiable?	Yes	No
Is the research question original? (i.e. *not appear on the first results page of google scholar?)*	Yes	No
Does the author cite at least 4 sources?	Yes	No
Are sources relevant to the research question?	Yes	No
Does the author point to a gap in the literature and explain how the research will address that gap?	Yes	No
Does the hypothesis express a directional Different based on a specific measure?	Yes	No

Comments on Introduction Section:

Data/Methods Section:

Does the experimental design correctly use a control group and randomization to minimize bias?	Yes	No
Can the experimental measure support or refute the hypothesis?	Yes	No

Comments for Data/Methods Section:

Discussion Section:

Does the author address potential confounds and limitations of the study?	Yes	No
Does the author state improvements for the study?	Yes	No
Does the author state implications for future study?	Yes	No

Comments for Discussion Section:

For the Score manipulation:
Attached are two papers for you to review.
The first author's score is in the same half of the class as yours,
and the second author's score is in the other half of the class than your
Please fill out a separate rubric for each review.

For the Location manipulation:
Attached are two papers for you to review.
The first author sits on the same side of the room as you do,
and the second author sits on the other side of the room than you do
Please fill out a separate rubric for each review.

Figure 1. Left: Rubric given to subjects in the Location Condition. Top right: Email sent to students in the Score Condition. Bottom right: Email sent to students in the location condition.

3.2.2 Distribution, Completion, and Evaluation of Rubrics. Students received an email containing links to two copies of the rubric in Google Forms. The rubric consisted of 11 yes/no questions in three sections: the introduction, methods, and discussion sections of the students' project proposals. After each section, room for optional comments on that section was provided. To draw attention to grouping, the rubric first asked whether or not the author was *Ingroup* or *Outgroup*. The rubric also asked whether the review was completed first or second, to control for fatigue caused by order of review (Figure 1). Students had three days to review submissions. Reviews were then sent back to the submitters. The authors rated the review quality on a scale of one to five stars.

3.3 Measures

3.3.1 Quantitative Variables: Time Spent and Review Length. Two quantitative variables measured outcome: time spent on review, and total number of characters in the optional comments. We normalized the time-spent data by log transforming the number of seconds that were spent on each review. However, time-spent data can be a noisy signal [6]. To account for students potentially walking away from their computer while completing the review, we excluded times that were three standard deviations above the mean.

We chose Review Length as a second quantitative measure of resource allocation. Because our data was non-normal, we modeled Character Count with a Poisson distribution, specifying a log-link. This distribution is commonly used with count data, and appropriate when a large number of responses are zero [35].

3.3.2 Qualitative Variables: Coding of Comments. Effort was qualitatively measured by the frequency of three coded comments types: "I like" (complement), "I wish" (critique), and "Here's How" (suggestion) [16,26]. "I like" comments consisted of positive remarks on the paper. "I wish" comments critiqued the paper, without offering a solution. "Here's how" comments offered an alternative. “Here’s how” comments represent the highest level of effort allocated by the reviewer because the reviewer must go beyond commenting and add their own contribution. Comments were coded separately by two independent reviewers and the average was used for analysis.

3.3.3 Other Measures. Reviewing harshness was measured by the score assigned to the submissions (how many rubric items were graded “yes,” out of 11 rubric items), and perception of review was measured by authors’ ratings (out of 5 stars) of the review they received.

3.3.4 Hypotheses for measures. We hypothesized that subjects who were *Ingroup* would spend more time and write more in both conditions, indicating higher effort allocation.

We planned to use the coding of comments to gain a more detailed sense of the attitudes that drove effort allocation, although our hypotheses were mainly exploratory and intended to guide future studies. We hypothesized that being in the same location would create a cooperative mindset, and thus subjects may include more “I like” comments to people in their group. Additionally, Score may elicit a competitive mindset—and thus students in the Score Condition might not be as helpful to those who need it. If this were the case, they would give more unsubstantiated negative feedback (more “I wish”, fewer “Here’s how”) than students in the Location Condition.

3.4 Data Analysis: Overall Model

Our analysis used a Generalized Linear Mixed Model because it is beneficial for analyzing data with repeated measures and potentially large individual differences [14]. The GLMM framework comprises fixed effects (predictors and factors), and random effects (slope and intercept of each subject). While the fixed effects represent the explanatory variables responsible for systematic variation in responses, the random effects allow the differences between individuals to be assessed [22]. The GLMM was executed by the linear modeling package lme4 made for the programming language R [2].

Our model analyzed the main effects and interaction of both Condition (Location vs. Score) and Review Type (*Ingroup*/*Outgroup*) on six outcome variables. Our main quantitative outcome variables were Time Spent (in seconds) and Review Length (in characters). We ran the same model on score on submission, rating of feedback and number of coded comments--I like, I wish, and Here’s how, separately.

Also included in the model were the variables that we wished to control for. The following were included as factors (controls): Reviewer Score (reviewer’s current class score), Author Score (author’s current class score), location of seat in the classroom (right or left side) Order of Review (whether or not the submission was reviewed first or second), and Submission Score (how many rubric items out of 11 were marked correct). As random effects, we had intercepts for subjects and items, as well as by-subject and by-item random slopes for the effects of Condition and Review Type. We first included the interaction of Condition and Review Type in every model and removed them from the model when they did not show significance [14,15].

Table 1 provides a summary of the significant and non-significant effects found. Included in the table are three coefficients from the GLMM output—b, SE, and the P-value. B is the point estimate—estimating the change in the dependent variable for every unit-change in the independent variable. B can be likened to a difference in means of the variable. SE represents the standard deviation of the point estimate, an indicator of precision/certainty about the estimate. The P-value reflects whether or not the point estimate has been calculated precisely enough to distinguish it from zero [14].

Study 1 Mixed Model Results

	Dependent variable:				
	Time Spent (seconds)	Review Length (characters)	I Like	I Wish	Here's How
	linear mixed-effects	*generalized linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*
	(1)	(2)	(3)	(4)	(5)
Condition (Score)	-0.154* (0.078)	-0.734* (0.344)	-0.227 (0.221)	-0.129 (0.304)	-0.038 (0.238)
Group (Outgroup)	-0.011 (0.030)	-0.074*** (0.011)	-0.134 (0.113)	0.115 (0.159)	0.110 (0.128)
Score on Submission	-0.013 (0.011)	-0.026* (0.010)	0.143*** (0.036)	-0.214*** (0.054)	0.026 (0.042)
Reviewer Grade	0.002 (0.006)	0.011 (0.026)	-0.049** (0.017)	0.069** (0.022)	0.055** (0.018)
Author Grade	0.003 (0.004)	-0.008 (0.010)	-0.001 (0.012)	-0.026 (0.019)	-0.033* (0.014)
Order of Review (first)	0.020 (0.035)	0.224*** (0.022)	-0.078 (0.120)	0.015 (0.182)	0.355* (0.142)
Location (right)	-0.116 (0.072)		0.211 (0.206)	0.104 (0.260)	0.268 (0.211)
Category X Group		0.243*** (0.036)			
Constant	2.880*** (0.307)	6.119*** (1.173)	2.053* (0.911)	2.512* (1.276)	-0.478 (1.002)
Observations	195	195	195	195	195
Note:					*p<0.05; **p<0.01; ***p<0.001

Table 1. Summary of Predictors and Factors for Quantitative and Qualitative Measurements in Experiment 1

4 RESULTS

We analyzed whether *Ingroup/Outgroup* student in the Location and Score conditions spent more time and wrote more on their reviews. In summary, students in the Location Condition spent more time and wrote longer reviews than students in the Score Condition. Contrary to our hypotheses, *Ingroup* students spent the same amount of time on reviews as *Outgroup* students, and *Ingroup* students actually wrote less than *Outgroup* students, when grouped by Score.

4.1 Time Spent

4.1.1 Main effect: students grouped by Score spend less time on reviews. We performed a GLMM on the Time Spent (in seconds) on reviews (M =1125, SD =1329), including Condition and Review Type as predictors while controlling for Reviewer Grade, Author Grade, Order of Review and Score on Submission. Condition showed a significant effect on Time Spent (b=−.154, SE=.078, p<.05), such that students who were grouped by score spent an average of 15% fewer seconds reviewing their peers [38] than did students who were grouped by location (Figure 2), all other factors remaining constant. Review Type did not predict time spent on review (b=−.011, SE=.030, p=.712), and there was no significant interaction between Condition and Review Type (b=−.022, SE=.059, p=.711).

4.1.2 Other predictors of Time Spent. Reviewer Grade in the class (b=.002, SE=.006, p=.699), Author Grade in the class (b=.003, SE=.004, p=.418), Order of Review (b=.020, SE=.035, p=.561), and Score on Submission (b=−.013, SE=.011, p=.239), were not predictors of Time Spent on review.

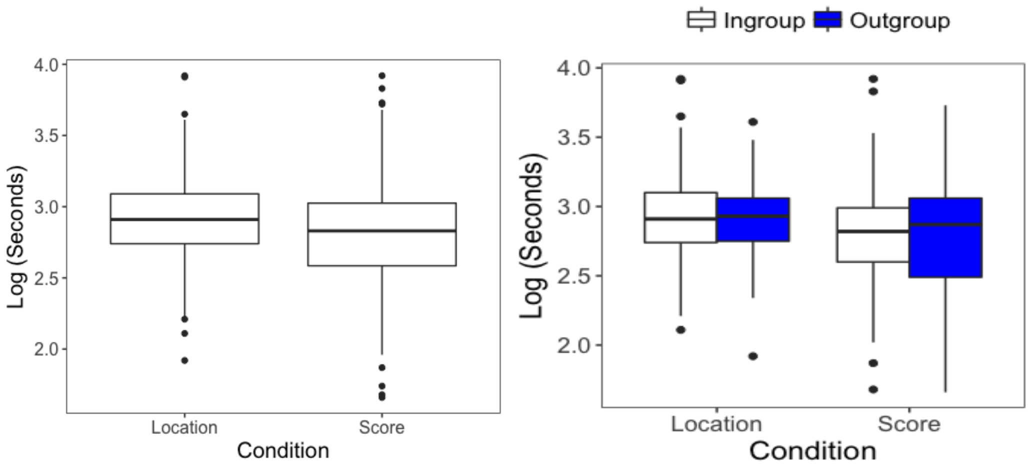

Figure 2. Students in the Score Condition spend less time on reviews than students in the Location Condition (measured in seconds)

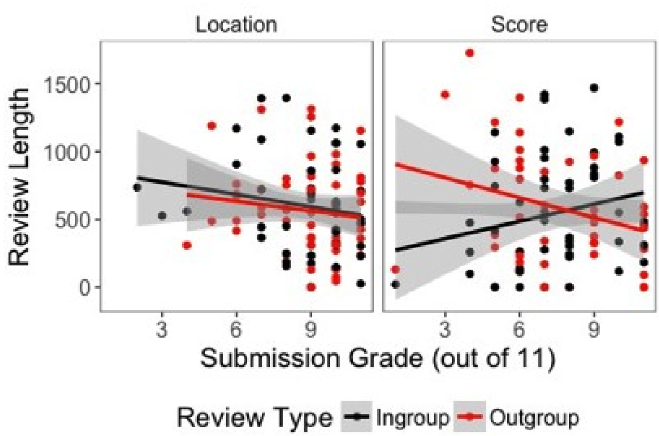

Figure 3. When grouped by Course Score, *Ingroup* students wrote shorter reviews than *Outgroup* students.

4.2 Review Length

4.2.1 Main effect: Students grouped by Score write less on Ingroup reviews. We performed a similar GLMM on the length (in characters) of each review (M = 581, SD = 385). Again we included Condition and Review Type as our independent variables, while controlling for Reviewer Grade, Author Grade, Order of Review and Score on Submission.

Condition had a significant effect on Review Length (b=−.734, SE=.344, p<.05), such that students in the Location Condition wrote longer reviews (Figure 3). Review Type also had a significant effect on total characters used (b= −.074, SE=.011, p<.001), but only in the Score Condition. This is characterized by an interaction between the two main effects, such that within the Score Condition, *Ingroup* students wrote significantly less on reviews than *Outgroup* students (b=.243, SE=.036, p<.001), all other factors remaining constant (Figure 3).

4.2.2 Other predictors of Review Length. Grade on Submission was also a significant predictor of Review Length, such that highly graded submissions (submissions with a higher number of rubric items marked correct) get less feedback than lowly graded submissions (b=−.026, SE=.011, p<.05). Order of Review was also a significant predictor of Review Length, such that reviews that were completed first were longer (b=.224, SE=.022, p<.001. Neither Reviewer Grade in the class (b=.011, SE=.026, p=.668), nor Author Grade(b=−.008, SE=.010, p=.398), in the class was a significant predictor of Review Length. See Supplementary Figure 4.

4.2.3 Control Analysis for Review Length. We considered, and then ruled out, the possibility that the effect of Review Type on Review Length was due to the experimental design of the Score Condition. The Score Condition was divided such that, for *Ingroup* students, weaker students would grade weaker students and stronger students would grade stronger students, while for Outgroup students, stronger students would grade weaker students and weaker students would grade stronger students. Because there may have been systematic differences in the review length of stronger students and weaker students, and in stronger reviews and weaker reviews [26], we could have expected differences in amount of feedback written for *Ingroup* students than *Outgroup*.

To understand whether or not this effect was due to the division of student ability in the Score Condition, we divided students in the Location Condition by the same low and high score criteria that was used to divide the students in the Score Condition. We then compared when stronger students graded stronger students and weaker students graded weaker students (as in the *Ingroup* Score Condition) to when stronger students graded weaker students and weaker students graded stronger students (as in the *Outgroup* Score Condition). We did not find the same result that we found in the Score Condition—namely, that *Ingroup* students wrote less on reviews. This suggests that the experimental design of the Score Condition was not the contributor to the effect (Figure 4).

4.3 Qualitative Analysis: Coded Comments

To gain a more nuanced understanding of effort spent, we coded the optional comments into three

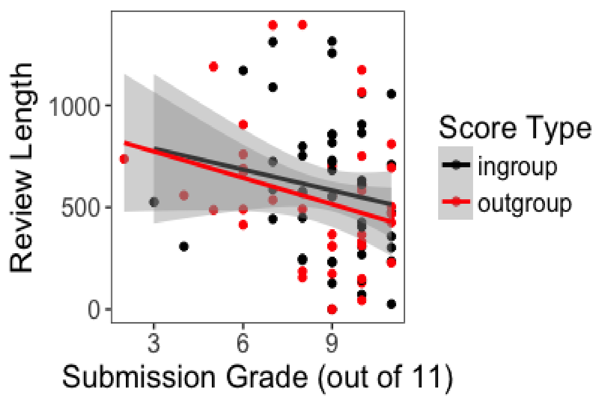

Figure 4. Students in the Location Condition grouped by score do not show an *Ingroup/Outgroup* effect on Review Length

categories—complimentary ("I like"), critical ("I wish") and suggestive ("Here's how"). We did not find any main effects of Condition or Review Type on comments, however we found that factors such as grade in the class predicted types of comments.

4.3.1 No main effects on complements, critiques, or suggestions. Neither Condition (b=-227, SE=.221, p=.296), nor Review Type (b=−.14, SE=.12, p=.231), significantly affected the total number of complementary ("I like…") comments (M = 1.14, SD = 1.27). Neither Condition (b=−.129, SE=.297, p=.664), nor Review Type (b=. 115, SE=.156, p=.466), had a significant effect on number of critical ("I wish") comments (M = 2.49, SD = 1.65). Finally, Neither Condition (b=−.037, SE=.233, p=.872), nor Review Type (b=.110, SE=.126, p=.386), was significantly predictive of how many high quality suggestions ("Here's how…") a reviewer included in their review (M = .93, SD = 1.32).

4.3.2 Better students give more suggestions. Reviewer Grade, Author Grade and Order of Review all predicted number of high quality suggestions in Study 1. Reviewers who had a higher current grade in the class were more likely to give high quality suggestions (here's how) in their reviews (b=.055, SE=.017, p<.01). Authors with a lower current grade in the class were more likely to receive more high quality suggestions than authors with a high current score in the class (b=−.033, SE=.014, p<.01). Unsurprisingly, students were more likely to include high quality suggestions if the review was completed first (b=.354, SE=.140, p<.05).

4.3.3 Better students give fewer complements. Reviewer Grade in the class was also a significant predictor of amount of positive feedback, as reviewers with a higher grade in the class wrote significantly fewer complementary comments (b= − .049, SE=.017, p<.001). Submission grade was a significant predictor, such that higher graded submissions received more positive feedback (b=.143, SE=.036, p<.001).

4.3.4 Better students give more critique, and better submissions get less critique. Reviewer Score in the class and Author Score in the class were significant predictors of peer critique. Reviewers with a higher current grade in the class critiqued more often (b=.069, SE=.022, p<.01), and authors with higher-scoring submissions received fewer "I wish" comments (b= −.17, SE=.05, p<.001).

4.4 Score on Submission

4.4.1 Main effect: Students grouped by Score give lower grades. To understand biases in grading, we modeled predictors of the total score assigned to the submissions, or number of rubric items out of 11 marked correct (M = 8.09, SD = 2.22). Condition was a significant predictor of assigned score (Figure 5), as students grouped by location gave one another significantly higher scores than students grouped by score in the class (b = -1.138, SE = .435, p ≈ .010). Again, no significant In/*Outgroup* difference was found (b = −.080, SE = .287, p = .781).

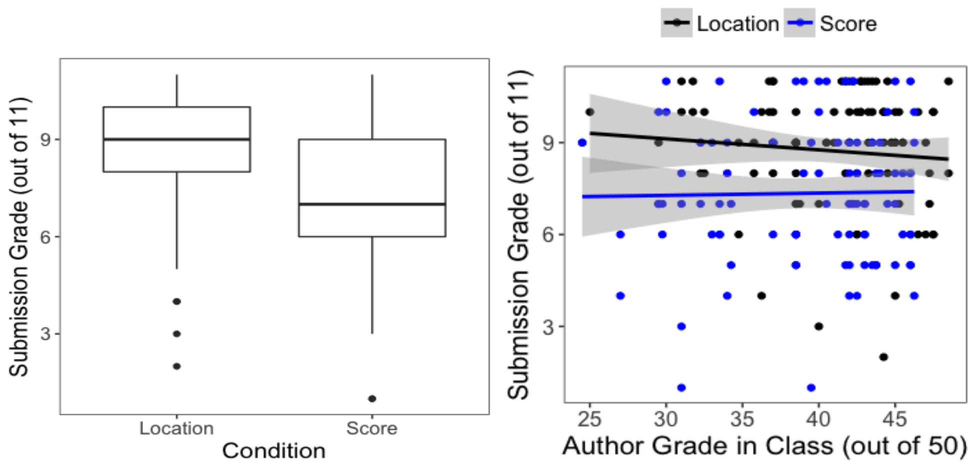

Figure 5. Students grouped by Location assigned one another higher scores than students grouped by Score.

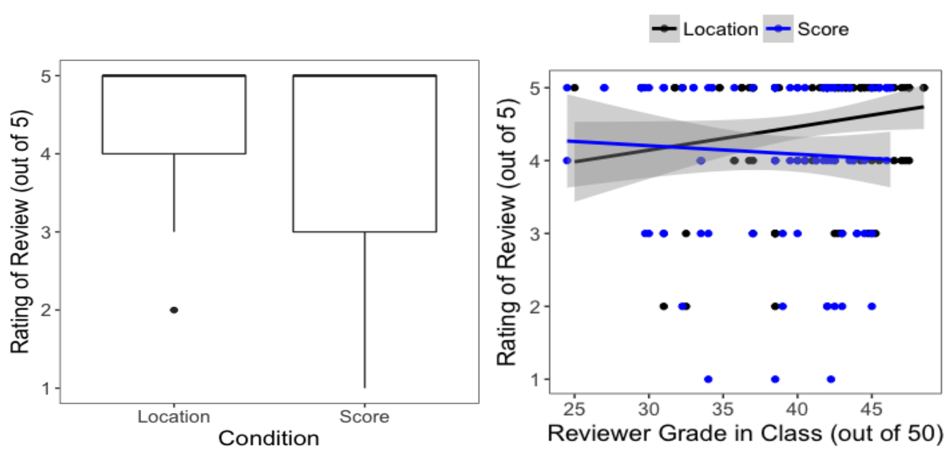

Figure 6. Students grouped by Score rated the quality of reviews lower

4.4.2 Better students give lower grades. The current grade in the class of the reviewer was also a predictor of assigned score on submission, such that reviewers with higher grades in the class gave lower scores (b = −.052, SE = .022, $p < .05$). The current score of the author was also a significant predictor of score on submission, such that students with a higher current grade in the class were given higher scores on the submission (b = .073, SE = .030, $p < .05$).

4.5 Perception of Reviews

4.5.1 Main effect: Students grouped by Score rate quality of reviews lower. After the reviews were returned to the original authors, the authors were asked to rate the quality of the review on a scale from 1 to 5 stars (M = 4.29, SD = 1.0). Condition (Figure 6) was a significant predictor of review ratings, such that students in the Score Condition rated the quality of their reviews significantly lower than did those in the Location Condition ($b=-.439$, $SE=.172$, $p<.05$). There was no significant effect of Review Type on Review Rating ($b=-.161$, $SE=.102$, $p=.122$).

4.5.2 Other effects. Students with a higher current course grade rated the quality of the review they received lower ($b=-.024$, $SE=.012$, $p<.05$).

5 CONCLUSION: STUDY 1

In Study 1, students in the Location Condition spent more time and wrote longer reviews than students in the Score Condition. Contrary to our hypotheses, *Ingroup* students spent the same amount of time on reviews as *Outgroup* students, and *Ingroup* students actually wrote less than *Outgroup* students. This effect was driven by an interaction between Condition and Review Type, such that *Ingroup* students in the Score Condition (but not the Location Condition) wrote less than *Outgroup* students. Condition also had a significant effect on peer grading, such that students in the Score Condition graded their peers lower than students in the Location Condition. We found no effects of Condition or Review Type on total amount of positive, negative, or constructive types of comments. Students in the Location Condition also rated the quality of reviews significantly higher than those in the Score Condition, suggesting that the students grouped by score were less satisfied with the quality of the reviews.

We hypothesized that students in the same group would spend more effort on *Ingroup* peer reviews, regardless of whether they were divided by Score or Location in the class. Location did not induce an *Ingroup* effect, and Score induced the reverse of an *Ingroup* effect. Although there was no *Ingroup*/*Outgroup* effect in the Location Condition, students grouped by Location spent more effort on their reviews that students grouped by Score. This led us to wonder whether attention to shared location in space might have motivated an increase in effort spent. Alternatively, drawing attention to Score might have caused an environment of competition, which may have decreased resource allocation. To address these outstanding questions, we performed a second study.

6 STUDY 2: WAS THE MINIMAL GROUP MINIMAL?

The results from Study 1 left outstanding questions, which we addressed in Study 2. Firstly, since students in the Location Condition spent more effort on reviews than students in the Score Condition, was location a stimulant or was score a depressant? A large body of work suggests that perceived location may enhance group feeling [3]. Sharing details of locations through photos of others' contexts increased resource allocation towards members of other groups [24]. It is possible that in our first study, drawing attention to shared location in space (the classroom) may have decreased social distance and caused the main effect of Condition. We tested this in the second study by adding a control group.

Secondly, was the reverse-*Ingroup* effect in the Score Condition dependent on the presence of a salient *Outgroup*? In the second study, students were only given essays from *Ingroup* authors, in order to assess whether or not a salient *Outgroup* was needed. Thirdly, a Motivational Condition was added, to assess whether or not group effects could be explicitly manipulated. This decision stems from prior research on successful motivational cues and belonging to a group. From research done on motivational effects of cues of working together [5,41,42], we hypothesized that this group would show an increase in effort spent on reviews.

6.1 Hypotheses Study 2

1. Students in the Location Condition would spend more effort on reviews than students in the Control Condition.

2. Students in the Score Condition would spend less effort on reviews than students in the Control Condition. Students in the Motivational Condition would spend more effort on reviews than any other group.

7 METHODS

Study 2 implemented a Parallel Group Design with 4 conditions, 1 control condition and 3 experimental conditions. Study 2's methods paralleled Study 1, with the addition of two conditions—a Control Condition, in which students were told nothing about their peers, and a Motivational Condition, in which students were also told nothing, but read a script concerning the importance of their contribution to the group.

7.1 Participants

In an "Introduction to Ethnographic Methods in Psychology" course at a western US research university, 144 students (ages 18—22) took part in a voluntary peer review assignment of their final project, which required a design of an ethnographic experiment and a five-page written proposal.

7.2 Procedure

Subjects were randomly assigned to one of four conditions: Location (N = 37), Score (N = 36), Control (N = 33), and Motivational (N = 38). The performance of the students (their score in the class) was equally distributed between the four groups. Students in the Location Condition were told they were grouped by the location they usually sat in the class, students in the Score Condition were told that they were grouped by their performance in the class so far, students in the control group were not told anything about the authors they were evaluating, and students in the motivational condition all read a motivational script concerning the role of student peer assessment in overall group success (Figure 7).

As in Study 1, students submitted a first draft of their final project through Google Forms, along with the side of the class they usually sit in and their current score in the class. Each student was given two Ingroup submissions to review. This varied from Study 1, in which each student was given one *Ingroup* and one *Outgroup* submission to review. We took this approach for two reasons. 1) Our main concern was to measure the difference between disclosing information about Location and disclosing information about Score, and 2) we wanted to understand whether the results would be affected if there was no *Outgroup*.

Students received links to two submissions to review, as well as copies of a rubric. The rubric was structured similarly to the rubric in Study 1, consisting of 11 yes/no questions and three comments sections. Its content was modified slightly to suit the ethnographic project proposal assignment. Students filled out the rubrics, which were then sent to the original authors, who rated the quality of the review from 1-5 stars. We measured differences in review quality by analyzing Review Length (in characters), types of comments made, and score given to the submissions. Due to a technical issue, Time Spent information is unavailable for Study 2.

Study 2 Linear Model Results

	Dependent variable:				
	Review Length (characters)	Score on Submission	I Like	I Wish	Here's How
	(1)	(2)	(3)	(4)	(5)
Condition (Location)	-51.159 (56.565)	-0.063 (0.331)	-0.060 (0.162)	-0.052 (0.275)	0.085 (0.147)
Condition (Motivation)	107.142 (59.429)	0.081 (0.348)	0.012 (0.170)	0.207 (0.289)	0.108 (0.154)
Condition (Score)	50.533 (57.524)	-1.003^{**} (0.331)	-0.124 (0.165)	-0.128 (0.280)	0.068 (0.149)
Score on Submission	-14.512 (10.909)		0.202^{***} (0.031)	-0.202^{***} (0.053)	-0.051 (0.028)
Reviewer Grade	0.937 (3.592)	-0.002 (0.021)	0.023^{*} (0.010)	0.0003 (0.017)	-0.012 (0.009)
Location (right)	10.710 (41.295)	0.030 (0.242)	-0.132 (0.118)	-0.143 (0.201)	0.112 (0.107)
Constant	639.426^{***} (189.408)	7.380^{***} (1.004)	-1.050 (0.542)	4.567^{***} (0.922)	1.333^{**} (0.491)
Observations	251	251	251	251	251
Note:			*p<0.05; **p<0.01; ***p<0.001		

Table 2. Quantitative measurements of review quality: Condition was not a predictor of Review Length. Students in the Score condition graded their peers lower. Qualitative measurements of review quality: I like, I wish, and Here's how are attributes of good feedback to assess review quality. There were no significant differences across conditions. Higher scored submissions received more I like feedback; Lower scored submissions received more I wish feedback.

8 RESULTS

In Study 2, we sought to understand what drove the main effects of study 1--namely whether attention to Location was a motivator or attention to Score was a de-motivator, compared to a Control Group and Explicitly Motivated group. We also tested whether or not the *Ingroup* Score bias found in Study 1 was present sans *Outgroup*. We found that without an *Outgroup*, there were no significant effects of Condition on Review Length, comment type, or Rating of Reviews. However, as in Study 1, students in the Score Condition graded their peers significantly lower than students in the other conditions. This indicates that the Score Condition functioned as a de-motivator, possibly due to the creation of a competitive environment.

8.1 Review Length

8.1.1 Main effect: no effect of Condition on Review Length. We performed a Linear Model measuring effects of group membership on Review Length, measured by character count (M = 512, SD = 362). Because each subject completed two *Ingroup* reviews, we used the average character count across both reviews for each subject. This served to normalize the data, as well as allow the use of a simpler linear model. No significant differences were found in Review Length between subjects in the four conditions. Compared to the Control condition, Location

Order of Review (is this the first or second submission you have reviewed?)______

Rubric Instructions:
Please answer the questions (highlight or underline yes or no) and give comments/suggestions below where appropriate. You may write as much in the comments section as you want. *Remember, you are all in this class together. Every bit of effort you put into giving feedback ends up benefiting the grades of the class as a whole.*

Figure 7. Rubric instructions for Motivation Condition

(b= -51.159, SE= 56.565, p=.367), Score (b= 50. 533, SE= .57.524, p=.381), and Motivational (b= 107. 142, SE= .59.429, p=.073) Conditions did not show any significant differences in Review Length.

8.2 Coded Comments

8.2.1 No main effects: grouping feature does not affect feedback quality. We performed three General Linear Mixed Models measuring effects of group feature on number of positive (I like…), negative (I wish…), and constructive (Here's how…) pieces of feedback. Condition had no significant effect on number of positive (M = 1.33, SD = 1.18), negative (M = 3.04, SD = 1.72), or constructive (M = .54, SD = .99) comments. See Table 2.

8.2.2 Better submissions received more complements and fewer critiques. Score of submission predicted amount of positive and negative feedback, but not constructive feedback. Higher graded submissions received more I likes (b= .202, SE= .031, p< .001), and fewer I wishes (b= − .202, SE= .053, p <.001). Students with a higher grade in the class gave more I likes (b= .023, SE=.010, p <.05).

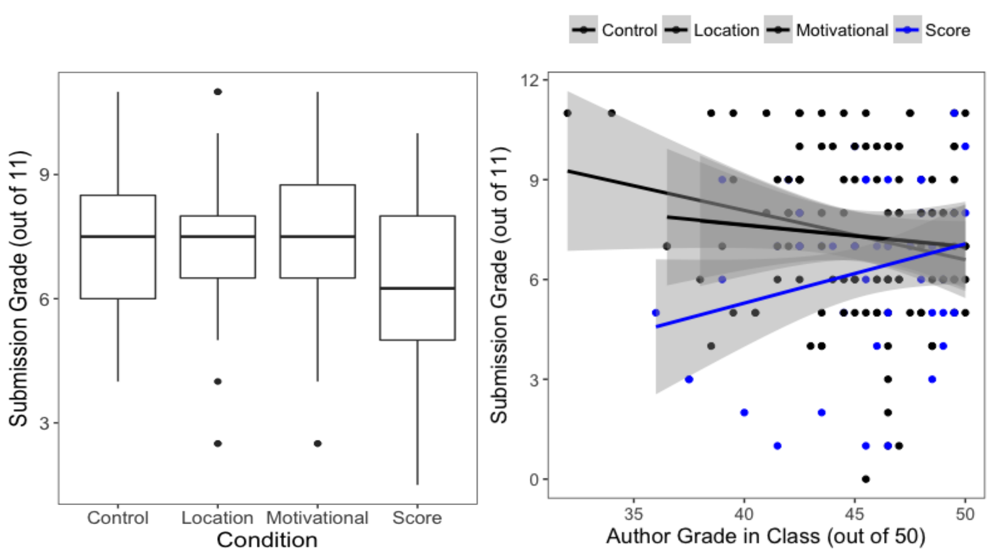

Figure 8. (left) Students in the Score Condition graded peers lower than students in the other conditions. (right) Students in the Score Condition gave grades that more closely tracked the author's course grade.

8.3 Score on Submission

8.3.1 Students in the Score Condition grade their peers lower. Condition had a strong effect on Submission Score (M =7, SD = 2.5) (Figure 8). As in Study 1, students in the Score Condition graded their peers significantly lower than did students in the Control condition (b=-1.003, SE= .331, p< .01).

8.3.2 *Worse students got higher scores on submissions (except in the Score Condition).* Author Grade in the class also significantly predicted Score on Submission, such that better students got significantly lower scores on their submissions. Students in the Score Condition gave grades that more closely tracked the author's current grade in the class. In other words, the Score Condition, peer-reviewed grades more closely reflected the teaching staff's assessment of current performance of the students overall. This last result will be elaborated on in the discussion.

8.4 Perception of Reviews

We performed a GLMM measuring the effects of Condition on Rating of Review (1-5 stars) (M = 4.45, SD = .90). We found no significant effects of Condition on review ratings.

9 CONCLUSION: STUDY 2

In Study 2, we found no significant effects of Condition on Review Length, feedback quality, or rating of reviews. As in Study 1, stronger students give more feedback and higher- graded submissions garner less feedback. Most significantly, as in Study 1, Condition was a strong predictor of assigned grades, with students in the Score Condition grading their peers significantly lower than students in the Control Condition.

10 DISCUSSION

Peer review, an important aspect of online learning, currently lacks the ability to motivate learners to spend time and cognitive effort grading their peers' submissions. This is potentially detrimental because learners in online courses lack the cohesion and group feeling that comes with in- class learning. This project investigated techniques for decreasing social distance between peers and intrinsically motivating learners to allocate effort to peer review.

According to recent research, social distance may be reduced through the formation of a common group identity that emphasized similarity [13]. For example, previous work on fostering commitment in online groups shows that simply having a team name, team logo, and shared team goal increased contribution to an online movie site [13].

In two peer review studies, we grouped students by Location and Score in the class, in the hopes of producing a bias toward allocation of resources to the *Ingroup*. Instead, we found that grouping students by location in the classroom is not an effective way to decrease social distance, and disclosing score can negatively impact how much students help their peers with perceived similar scores.

10.1 Explicit cues of group-ness did not affect effort

Notably, the condition in which students read a motivational script about group membership did not affect effort spent or quality of reviews. From research done on motivational effects of cues of working together [5,39,41], one would expect this group to have shown an increase in effort spent on reviews. This finding suggests that explicit reminders of being part of a group do not create a sense of belonging any more than reminders of manufactured common group features do.

10.2 Knowledge of Score is a complicated de-motivator

The lower grades in the Score Condition in both studies suggest that thinking about others' score in the class fosters a competitive environment in which students grade one another more harshly. While it could be argued that students in the Score Condition are critiquing their peers in order to be more helpful, evidence for the competitive environment argument stems from the relationship between grade and amount of feedback given in the comments sections. Although students in the Score Condition graded their peers significantly lower, they did not give a higher amount of feedback in the comment sections to help clarify these low grades.

It is likely that the decrease in effort spent on *Ingroup* reviews in the Score Condition in Study 1 was due to this competitive environment created by the Score Condition, and suggests that students offer less help to potential competitors. Importantly, Study 2, which had students review two *Ingroup* reviews, did not show a significant effect. This suggests that the presence of an *Outgroup* is necessary for the negative effect of the *Ingroup* in the Score Condition.

10.3 Advantages of a critical mindset

Interestingly, we also found a negative relationship between Author Score and Grade on Submission. We looked into this, and found that every Condition except the Score Condition graded lower-ability students as high as, if not higher than, higher-ability students (Figure 8).
This suggests that although students in the Score Condition did not help lower ability students any more than those in other conditions, knowledge of others' score may invite a critical mindset, which decreases inflationary grading. In the other conditions, students did not give low grades, potentially because they were "too nice" to their fellow students. In the Score Condition, peer-reviewed grades more closely reflected the teaching staff's assessment of performance of the students overall, as measured by their current grade in the class. (Figure 8). It is possible that a critical mindset might be helpful in certain circumstances, when honesty in feedback is important.

10.4 Is the Location manipulation a nominal group?

The Location grouping was meant as a nominal assignment, measuring any effect of forming groups that are neither meaningful nor related to the task. We hypothesized that any *Ingroup*/*Outgroup* difference in response would indicate that minimal group theory applies to peer review. We found no *Ingroup*/*Outgroup* difference in the Location Condition, but we did find an unexpected main effect of Location over Score. The Location grouping emphasizes a shared physical space, whether *Ingroup* or *Outgroup*, making the group more substantive than nominal. Study 2 investigated this by adding other types of potentially nominal groups, and found that student response to a Location grouping was similar to that of Control reviewer that were not assigned to a group. It is possible--even likely—that a more substantive Location grouping (e.g., students at rival schools, or residents of different countries) would elicit *Ingroup*/*Outgroup* effects.

10.5 Design Implications

This project made headway in understanding the effect of disclosing features on effort spent helping peers. Although we hypothesized, based on social-psychological research, that an *Ingroup*/*Outgroup* effect may be easily elicited with any feature of similarity, this turned out not to be the case. Features matter, and our results highlight the complicated effects of being grouped by Score.

The fact that we found similar effects in two different classes suggests that they might be generalizable to peer-review in a classroom/course setting. In our study, although students took part in a physical class, students reviewed their peers anonymously online. It remains to be tested how this would compare to peer reviews in a fully online classroom, or even other online forums such as stack exchange or Wikipedia, where formative feedback is encouraged. In a completely online setting, where location in space is not shared, emphasis on location would likely have an effect.

Ultimately, if disclosing features can reduce social distance between peers in reciprocal social computing tasks like peer review, then easy interventions may be put in place to increase learner-centered collaboration in online environments, where the goal is large-scale, high-quality assessment.

REFERENCES

[1] Daniel Balliet, Junhui Wu, and Carsten K. W. De Dreu. 2014. Ingroup favoritism in cooperation: A meta-analysis. *Psychol. Bull.* 140, 6 (2014), 1556–1581. DOI:https://doi.org/10.1037/a0037737

[2] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw. Vol 1 Issue 1 2015* (2015). DOI:https://doi.org/10.18637/jss.v067.i01

[3] Erin Bradner and Gloria Mark. 2002. Why Distance Matters: Effects on Cooperation, Persuasion and Deception. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (CSCW '02), 226–235. DOI:https://doi.org/10.1145/587078.587110

[4] Marilynn B. Brewer. 1979. In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychol. Bull.* 86, 2 (1979), 307–324. DOI:https://doi.org/10.1037/0033-2909.86.2.307

[5] Priyanka B. Carr and Gregory M. Walton. 2014. Cues of working together fuel intrinsic motivation. *J. Exp. Soc. Psychol.* 53, (2014), 169–184. DOI:https://doi.org/10.1016/j.jesp.2014.03.015

[6] Justin Cheng, Jaime Teevan, and Michael S. Bernstein. 2015. Measuring Crowdsourcing Effort with Error-Time Curves. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 1365–1374. DOI:https://doi.org/10.1145/2702123.2702145

[7] Winnie Cheng and Martin Warren. 1997. Having second thoughts: Student perceptions before and after a peer assessment exercise. *Stud. High. Educ.* 22, 2 (January 1997), 233–239. DOI:https://doi.org/10.1080/03075079712331381064

[8] Kwangsu Cho and Christian D. Schunn. 2007. Scaffolded Writing and Rewriting in the Discipline: A Web-based Reciprocal Peer Review System. *Comput Educ* 48, 3 (April 2007), 409–426. DOI:https://doi.org/10.1016/j.compedu.2005.02.004

[9] Phil Davies. 2000. Computerized Peer Assessment. *Innov. Educ. Train. Int.* 37, 4 (January 2000), 346–355. DOI:https://doi.org/10.1080/135580000750052955

[10] Carol S. Dweck. 2008. *Mindset: The New Psychology of Success.* Ballantine Books, New York.

[11] Naomi Ellemers, S. Alexander Haslam, and Dick de Gilder. 2008. Motivating Individuals and Groups at Work in the 21st Century. In *21st Century Management: A Reference Handbook.* SAGE Publications, Inc., Thousand Oaks, II–182–II–192. Retrieved from http://sk.sagepub.com/reference/management/n68.xml

[12] Rosta Farzan, Laura A. Dabbish, Robert E. Kraut, and Tom Postmes. 2011. Increasing Commitment to Online Communities by Designing for Social Presence. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (CSCW '11), 321–330. DOI:https://doi.org/10.1145/1958824.1958874

[13] Dilrukshi Gamage, Mark E. Whiting, Thejan Rajapakshe, Haritha Thilakarathne, Indika Perera, and Shantha Fernando. 2017. Improving Assessment on MOOCs Through Peer Identification and Aligned Incentives. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (L@S '17), 315–318. DOI:https://doi.org/10.1145/3051457.3054013

[14] Andrew Gelman and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, Cambridge. Retrieved from https://www.cambridge.org/core/books/data-analysis-using-regression-and-multilevelhierarchical-models/32A29531C7FD730C3A68951A17C9D983

[15] Andrew Gelman, Jennifer Hill, and Masanao Yajima. 2012. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *J. Res. Educ. Eff.* 5, 2 (April 2012), 189–211. DOI:https://doi.org/10.1080/19345747.2011.618213

[16] Stephanie J. Hanrahan and Geoff Isaacs. 2001. Assessing Self- and Peer-assessment: The students' views. *High. Educ. Res. Dev.* 20, 1 (May 2001), 53–70. DOI:https://doi.org/10.1080/07294360123776

[17] Catherine M. Hicks, Vineet Pandey, C. Ailie Fraser, and Scott Klemmer. 2016. Framing Feedback: Choosing Review Environment Features That Support High Quality Peer Assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 458–469. DOI:https://doi.org/10.1145/2858036.2858195

[18] Julia H. Kaufman and Christian D. Schunn. 2011. Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instr. Sci.* 39, 3 (2011), 387–406.

[19] Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale* (L@S '15), 75–84. DOI:https://doi.org/10.1145/2724660.2724670

[20] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and Self Assessment in Massive Online Classes. *ACM Trans Comput-Hum Interact* 20, 6 (December 2013), 33:1–33:31. DOI:https://doi.org/10.1145/2505057

[21] Kimberly Ling, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid, Paul Resnick, and Robert Kraut. 2005. Using social psychology to motivate

contributions to online communities. *J. Comput.-Mediat. Commun.* 10, 4 (January 2005). DOI:https://doi.org/10.1111/j.1083-6101.2005.tb00273.x

[22] Steson Lo and Sally Andrews. 2015. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6, (2015), 1171. DOI:https://doi.org/10.3389/fpsyg.2015.01171

[23] Yanxin Lu, Joe Warren, Christopher Jermaine, Swarat Chaudhuri, and Scott Rixner. 2015. Grading the Graders: Motivating Peer Graders in a MOOC. In *Proceedings of the 24th International Conference on World Wide Web* (WWW '15), 680–690. DOI:https://doi.org/10.1145/2736277.2741649

[24] Jennifer Marlow and Laura Dabbish. 2012. Designing Interventions to Reduce Psychological Distance in Globally Distributed Teams. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (CSCW '12), 163–166. DOI:https://doi.org/10.1145/2141512.2141568

[25] M. M. Nelson and C. D. Schunn. 2009. The nature of feedback: How different types of peer feedback affect writing performance. *Instr. Sci.* 37, (July 2009), 375–401.

[26] Melissa M. Patchan, Brandi Hawk, Christopher A. Stevens, and Christian D. Schunn. 2013. The effects of skill diversity on commenting and revisions. *Instr. Sci.* 41, 2 (2013), 381–405. DOI:https://doi.org/10.1007/s11251-012-9236-3

[27] Maria Perez-Ortiz and Rafal K. Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *ArXiv171203686 Cs Stat* (December 2017). Retrieved from http://arxiv.org/abs/1712.03686

[28] Ken Reily, Pam Ludford Finnerty, and Loren Terveen. 2009. Two Peers Are Better Than One: Aggregating Peer Reviews for Computing Assignments is Surprisingly Accurate. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (GROUP '09), 115–124. DOI:https://doi.org/10.1145/1531674.1531692

[29] Yuqing Ren, F. Maxwell Harper, Sara Drenner, Loren Terveen, Sara Kiesler, John Riedl, and Robert E. Kraut. 2012. Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *MIS Q* 36, 3 (September 2012), 841–864.

[30] Todd Rogers and Avi Feller. 2016. Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting. *Psychol. Sci.* 27, 3 (March 2016), 365–374. DOI:https://doi.org/10.1177/0956797615623770

[31] J. Sanchez-Burks, R. E. Nisbett, and O. Ybarra. 2000. Cultural styles, relational schemas, and prejudice against out-groups. *J. Pers. Soc. Psychol.* 79, 2 (August 2000), 174–189.

[32] Stanley Schacter. 1959. *The Psychology of Affiliation; Experimental Studies of the Sources of Gregariousness.* Stanford University Press, Stanford, California.

[33] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. 2006. Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 2 (February 2006), 70–76. DOI:https://doi.org/10.1016/j.tics.2005.12.009

[34] Garriy Shteynberg and Adam D. Galinsky. 2011. Implicit coordination: Sharing goals with similar others intensifies goal pursuit. *J. Exp. Soc. Psychol.* 47, 6 (November 2011), 1291–1294. DOI:https://doi.org/10.1016/j.jesp.2011.04.012

[35] Walter W. Stroup. 2013. *Generalized linear mixed models: modern concepts, methods and applications.* CRC Press, Boca Raton, Fl.

[36] Henri Tajfel. 1974. Social identity and intergroup behaviour. *Inf. Int. Soc. Sci. Counc.* 13, 2 (April 1974), 65–93. DOI:https://doi.org/10.1177/053901847401300204

[37] Yla R. Tausczik, Laura A. Dabbish, and Robert E. Kraut. 2014. Building Loyalty to Online Communities Through Bond and Identity-based Attachment to Sub-groups. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 146–157. DOI:https://doi.org/10.1145/2531602.2531688

[38] UCLA: Statistical Consulting Group. Introduction to SAS. Retrieved from https://stats.idre.ucla.edu/sas/faq/how-can-i-interpret-log-transformed-variables-in-terms-of-percent-change-in-linear-regression/

[39] Siri Vinther, Ole Haagen Nielsen, Jacob Rosenberg, Niels Keiding, and Torben V. Schroeder. 2012. Same review quality in open versus blinded peer review in "Ugeskrift for Læger." *Dan. Med. J.* 59, 8 (August 2012), A4479.

[40] Gregory M. Walton and Shannon T. Brady. The Many Questions of Belonging. In *Handbook of Competence and Motivation: Theory and Application* (2nd ed.), Andrew J. Elliot, Carol S. Dweck and Yeager (eds.). Guilford Press, New York.

[41] Gregory M. Walton and Geoffrey L. Cohen. 2011. Sharing motivation. In *Social motivation.* Psychology Press, New York, NY, US, 79–101.

[42] Gregory M. Walton, Geoffrey L. Cohen, David Cwir, and Steven J. Spencer. 2012. Mere belonging: the power of social connections. *J. Pers. Soc. Psychol.* 102, 3 (March 2012), 513–532. DOI:https://doi.org/10.1037/a0025731

Received April 2018; revised July 2018; accepted September 2018.