HIGHER EDUCATION RESEARCH & DEVELOPMENT
https://doi.org/10.1080/07294360.2021.1969541

Peer feedback and teacher feedback: a comparative study of revision effectiveness in writing instruction for EFL learners

Ying Cui [a], Christian D. Schunn [b] and Xiaosong Gai [a]

[a]School of Psychology, Northeast Normal University, Changchun, People's Republic of China; [b]Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

This study investigated the revision effectiveness of peer feedback in comparison with teacher feedback, before and after a peer feedback training intervention that was designed to be implementable in large teaching load contexts and with EFL students. Fifty-six EFL students across two different class sections received teacher or peer feedback before and after training across five different writing assignments, with feedback conditions manipulated within and between class sections. Results show that EFL peer reviewers can provide more meaning-focused feedback than do teachers and that the impact of each piece of feedback on revision quality improves after training.

ARTICLE HISTORY
Received 28 January 2021
Accepted 22 July 2021

KEYWORDS
Teacher feedback; peer feedback; training

Introduction

Teacher feedback on writing is the dominant mode for both English as a second language (ESL) and English as a foreign language (EFL) classrooms in higher education. Teacher feedback to English language learners is considered effective (Connor & Asenavage, 1994; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001; Leki, 1990; Ma, 2018; Yang et al., 2006), and ESL and EFL students generally prefer teacher feedback to other forms of feedback (Nelson & Murphy, 1993; Zhang, 1995; Zhao, 2010). However, teacher feedback on writing in the university context can involve a heavy workload across a large class or many smaller sections. In such common circumstances, it is unlikely that teachers can provide careful comments to every student; and hastily produced feedback from teachers (e.g., edits without explanations) can lead to passive use of feedback without fully understanding the feedback, which necessarily limits the learning benefit (Ho et al., 2020; Hyland, 1998; Lee, 2007; Zhao, 2010). Therefore, considerable attention in higher education has been given to peer feedback as an alternative or complementary strategy to teacher feedback. We define peer feedback as comments and suggestions given to students on their documents by other students in the same class.

Research on peer feedback has generally found it to be effective in terms of cognitive, social, and linguistic aspects in both L1 (first language) and L2 (second language) writing classrooms at secondary and tertiary levels (Berg, 1999; Cho & Schunn, 2006; Ghani &

CONTACT Xiaosong Gai gaixiaosong@126.com School of Psychology, Northeast Normal University, 5268 Renmin Road, Changchun, Jilin Province 130024, People's Republic of China

Asgher, 2012; Guerrero & Villamil, 1994; Huisman et al., 2019; Kroll & Vann, 1981; Leki, 1990; Paulus, 1999; Rollinson, 2005; Ruegg, 2015; Saito & Fujita, 2004; Schunn et al., 2016; Topping, 1998; Tsui & Ng, 2000; Villamil & Guerrero, 1998; Yang et al., 2006). Nonetheless, limitations also can be found in the practice of peer feedback (Adachi et al., 2018). For Asian students, the linguistic distance between L1 and English is often large, and many university students have relatively weak knowledge of English. This relatively weak grasp of English can influence the amount and type of feedback that peers produce in EFL/ESL classrooms (Allen & Mills, 2016; Huisman et al., 2017; Snowball & Mostert, 2013) and also lead students to hold negative attitudes toward peer feedback (Carson & Nelson, 1994; Nelson & Murphy, 1993; Zhang, 1995; Zou et al., 2018). A few studies, however, suggest that peer feedback can improve both the quantity and quality of text revisions after students are trained to provide peer feedback (Berg, 1999; Lam, 2010; Min, 2005, 2006; Stanley, 1992). Here, we test the relative benefits of teacher vs. trained and untrained peer feedback in an EFL context.

Teacher feedback vs. peer feedback in ESL and EFL contexts

Some studies testing the effects of teacher versus peer feedback on revision suggest that teacher feedback resulted in more revision (Connor & Asenavage, 1994; Paulus, 1999). For example, Connor and Asenavage (1994) investigated the impact of peer and teacher feedback on essay revisions of freshmen ESL students in a US university. They found that although a large number of revisions were made, only 5% of the revisions resulted from peer feedback. Similarly, Paulus' (1999) examination of the impact of peer and teacher feedback on document revisions with 11 ESL students at a US university revealed that peer feedback accounts for only 14% of revisions whereas teacher feedback accounts for 34%.

However, studies conducted in Asian contexts found a mixture of positive and negative impacts of peer feedback, particularly when considering meaning-related revision (i.e., meaning of the text changes), surface-related revision (meaning of the text does not change, such as with spelling and grammar), and motivational aspects. Yang et al. (2006) examined two groups of students at a Chinese university and found that teacher feedback accounted for more writing revisions, but peer feedback resulted in a greater degree of student autonomy. Ruegg (2015) compared the effects of peer and teacher feedback on improvements in EFL students' writing ability at a Japanese university. Teacher feedback led to more meaning-level changes and higher grammar scores, while peer feedback led to more improvements to organization and academic style. Finally, EFL students have concerns about their abilities to provide useful feedback, as a mixture of their low self-efficacy in English and a strong cultural focus on the teacher as the authority (Tsui & Ng, 2000; Zou et al., 2018). Training on how to provide useful feedback may be needed to address both of these concerns.

Peer feedback training with ESL and EFL learners

Some university students report no prior experience at all with peer feedback and others report no prior training with peer feedback (Boud & Molloy, 2013; Min, 2005; Zhao, 2010; Zhu, 1995; Zou et al., 2018). Thus, it is not surprising that best practice recommendations (e.g., Berg, 1999; Lam, 2010; Min, 2005, 2006; Ren & Hu, 2012) often suggest the

use of careful training, especially in EFL environments, to increase the effectiveness of peer feedback, along with combining written and face-to-face peer feedback to improve peer feedback quality (Asparo et al., in press; Schillings et al., 2020; Zhu & Carless, 2018). The general idea is to improve student feedback literacy, which includes appreciating, evaluating, and acting on received feedback while managing emotional aspects (Carless & Boud, 2018).

Berg (1999) developed and tested a peer feedback training approach that included 11 steps organized into three main components addressing the various aspects of peer feedback literacy. First, students were asked to provide some peer feedback. Second, the instructor demonstrated how to use peer feedback, and then students practiced making use of peer feedback. Third, the instructor and students discussed problems encountered in using peer feedback and ways to improve the effectiveness of peer feedback. Results of Berg's study investigating this training approach with ESL students at a US university indicated that trained peer feedback generated more meaning changes in the revised drafts and yielded higher writing quality scores.

Min (2005, 2006) tested a more labor-intensive training approach composed of two phases. The first phase involved teacher modeling that demonstrated how to clarify writers' intentions, identify the source of problems, explain the nature of problems, and make specific suggestions. The second phase, the labor-intensive step, involved teacher-students conferences, in which the teacher helped students better understand peer feedback and how to generate more specific comments by examining reviewers' comments. Two studies of this approach, conducted with EFL students at an intermediate proficiency level at a Taiwanese university, revealed that the extensive coaching approach had a number of positive benefits: the students were better able to accept peer reviews and generated more relevant peer feedback at the meaning level, resulting in higher quality revisions. Lam (2010) suggested a variation of Min's approach that added an exploration phase before the teacher-student conference phase, but there was no systematic study of this variation.

It is important to note that both of these approaches focused on meaning-related revision, likely since writing research and practice places a strong emphasis on this aspect of writing. However, meaning-related changes may not be of particular importance to EFL students or in the assignments they often give in EFL classes, given that EFL students are still heavily focused on learning the basics of English.

Purposes and research questions of the present study

Despite the existence of many studies in the field of ESL and EFL on peer vs. teacher feedback, no studies have systematically compared teacher feedback with peer feedback before and after training. Further, relatively few studies have compared teacher feedback with peer feedback or trained with untrained peer feedback in the context of students coming from non-alphabetical languages. In the case of non-alphabetical L1, there can be a very large number of problems to correct, which may be overwhelming to teachers given the volume of students. Having peer feedback from multiple peers, who each have a much lower feedback load, might produce feedback that addresses more problems. At the same time, peers may be poorly positioned to correctly diagnose the problems, particularly the surface-related problems given a relative lack of familiarity with English.

While peer feedback training has been successfully tested in EFL contexts, the approach successfully used in the non-alphabetical context was labor intensive, and thus difficult to scale in many universities where instructors have a high course load. For example, in Asia, English writing instructors often have multiple sections of 50 students in a given semester (e.g., Ren & Hu, 2012; Zou et al., 2018). Having student-teacher conferences with all students is simply not feasible. A more scalable training approach needs to be tested in this context.

This study addressed three research questions related to peer feedback training involving an approach that could be more broadly scaled:

(1) Do teachers and peers with and without training provide different amounts and types of feedback at the meaning and surface levels?
(2) To what extent does each piece of teacher and peer feedback with and without training result in text revision?
(3) To what extent does overall teacher feedback or peer feedback with and without training result in higher quality text revisions?

Methods

Participants

The participants were 58 English majors (54 female; 4 male) who were juniors (mean age of 21) at a private university. All had received formal English training for more than eight years at the time of this study. However, their average score on the Test for English Majors-Band 4 (TEM-4) was only 57 out of 100, which is a relatively low score. They had not previously received training on peer feedback before the study. Two participants were excluded because they failed to submit papers, leaving fifty-six in the study.

Course setting

This study was conducted within two sections of an English writing course called 'Intermediate English Writing', offered for English majors by the English department. One section had 22 students and the other section had 34 students. The objective of the course is to develop writing skills in argumentation, although all of the students are still working to improve more fundamental aspects of English. The instructor, who was also one of the research team members, met the students once a week for 18 weeks, with each class session lasting 90 min. Students wrote five formal out-of-class papers during the semester, with each writing assignment involving a first draft completed as homework, feedback provided by either teacher or peer (depending on the assignment/condition) as comments on the essay printout, and finally a second draft completed as homework.

The specific writing assignments were taken from the textbook and involved argumentative writing on topics of general interest to university students: 'Spend or Save', 'Campus Love – Pros or Cons', 'Icon Worship', 'Examinations: For or Against', and 'Learning in School'. The class instruction and textbook emphasized a particular aspect of argumentation, such as organization, claim, refutation, emotional appeal, etc.

However, the basic writing task was the same across all five assignments and the same overall evaluation rubric was applied to each document.

Research design

The research design was a mixed between/within design across the two course sections as they were both sequentially transitioned from receiving teacher feedback to untrained peer feedback to trained peer feedback, but for different writing assignments. The teacher feedback was provided by another teacher in the same college who had previously taught this course several times. One section transitioned to untrained peer feedback for the second writing assignment and to trained feedback on the third writing assignment. The other section transitioned to untrained peer feedback for the fourth assignment and to trained peer feedback by the fifth writing assignment (see Figure 1). Within a section, the effects of different feedback source can be tested in the sequential transition from one feedback type to another (e.g., from teacher feedback to untrained peer feedback by comparing the first to second assignment in the accelerated section). Natural variation in students across sections is held constant in these comparisons. Between sections, specific assignments also allow for the contrast of condition effects (e.g., of teacher vs. untrained peer feedback within assignment 2). These within-assignment contrasts control for the effects of assignment topic and other temporal factors that naturally vary across the semester (e.g., instruction on argumentative writing, workload in other courses, overall experience with peer feedback). Consistency in patterns in the within-section and between-section contrasts provide a strong test of the condition effects. Performance on the first assignment is used to check for equivalence in student writing ability across sections.

Although the focus of this study is on the effects of receiving feedback from different sources, it is important to acknowledge that when there was peer feedback, students both received and provided peer feedback. Follow-up analyses assessed the effects of receiving feedback separately from the effects of providing feedback.

Peer feedback training

The training procedure was adapted from Berg (1999). The adaptations involved providing students clear guidance on effective peer feedback practices and some personalized

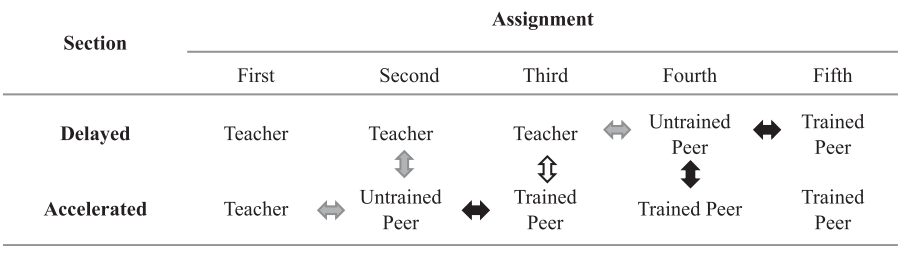

Figure 1. Sequence of feedback source across the five writing assignments within each of the two course sections. Arrows indicate tests of condition: teacher vs untrained in gray, teacher vs. trained in white, and untrained vs. trained peer in black.

feedback but using small groups rather than student-teacher conferencing to make it more scalable. There were three training sessions, and each of them lasted for one hour and a half. Throughout the training sessions, participants were divided into groups of three to four. In the first session, the instructor introduced peer feedback to participants and gave each student a feedback guidance sheet (see Appendix). Then, the instructor explained the items on the guidance sheet with illustrations. In the second session, the instructor selected three papers at different performance levels written by participants in previous assignments and displayed them in class. Then the instructor guided participants to give feedback on those papers. Throughout this guidance, the instructor encouraged participants to find problems at both the surface and meaning levels (Faigley & Witte, 1981) and then asked them to answer the questions about meaning-level problems in the guidance sheet. In the third session, the groups of students were given a paper written by one of their group members and asked to discuss ways of giving their feedback to the writer. In the course of the discussions, the instructor visited each group to answer questions. Note that throughout training and then later for the actual assignments, both teachers and peers provided marks/comments directly on the essays, rather than by filling out a reviewing form. This approach was taken to ensure that the teacher feedback, untrained peer feedback, and trained peer feedback conditions were equivalent in the physical way feedback was provided to peers. Further, since the majority of the essay problems were related to surface features, it would have been challenging for feedback givers to describe all of the problems via endnotes, rather than marks made on the essays themselves.

Measures

Writing quality. To study condition effects on writing quality improvements from feedback, the first and second drafts for each writing assignment were scored. The drafts were scored by two experts (English teachers) who were blind to condition and draft number, using rubrics from the TEM-4 involving: meaning, organization, and surface. Both graders had extensive prior experience using these rubrics. The inter-rater correlation in the overall draft score was high, $r = 0.71$, $p < 0.01$.

Amount of feedback. Each piece of feedback was coded for whether following the feedback would involve meaning or surface changes (i.e., whether or not it would change the text meaning) in the second drafts, following Faigley and Witte (1981). The amount of feedback was separately summed by type across feedback provided and received. This produced four values on each assignment for each student: # of surface comments received, # of meaning comments received, # of surface comments provided, and # of meaning comments provided. Note that the latter two were zero (and not included in analyses) in the case of teacher feedback conditions. Inter-rater reliability for these measures was almost perfect, with an average $r > .99$.

Analyses

The analyses were organized according to the research questions. First, the mean amount of feedback a student received on each writing assignment in each condition at the surface and meaning levels was examined using simple *t*-tests that focused on theoretical

contrasts highlighted in Figure 1. Between-condition comparisons within an assignment used unpaired t-tests and within-condition comparisons between assignments used paired *t*-tests. Cohen's *d* (difference in means divided by the average standard deviation) was used to measure effect size.

Second, the relationship of feedback amounts of different types and sources to revision behaviors was examined through regression analyses. The regressions included data from all the assignments that involved a given feedback source (e.g., all cases of teacher feedback) to increase statistical power. Dummy codes for the assignment were added to address possible confounds with assignment differences. Draft 2 score was the dependent variable, and draft 1 score along with various amounts of feedback experienced were used as predictors. An initial model exploration was done to determine which feedback variables predicted draft 2 quality (e.g., meaning vs. surface, provided vs. received). Partial η^2 was used to measure the effect size of a given predictor.

Third, the relationship of feedback source to final draft 2 quality was examined through regression models, using draft 1 quality as a covariate. Condition or assignment was used as a predictor depending upon the contrast. Partial η^2 was used to measure effect size.

Results

Teacher vs. untrained/trained peer feedback on amounts of meaning and surface feedback

Figure 2 presents the mean amount of surface and meaning feedback received. It also notes the size of the observed key contrasts from Figure 1 as a Cohen's *d*, along with their statistical significance level. Table 1 presents the means and standard deviations along with all the between-group statistical contrasts. At baseline in the first assignment, the two classes received an equivalent amount of surface comments and no meaning comments on their first drafts for the teacher as a feedback source.

Turning to the contrast of untrained peer vs. teacher feedback, the peers generally provided more surface comments than did the teacher with moderate-to-large effect sizes, and a small and sometimes non-significantly greater amount of meaning-related comments. In general, both the teacher and the untrained peers provided few meaning-related comments and a large number of surface comments.

After training, the peers somewhat reduced the amount of surface comments they produced, especially relative to the large amount provided on the second assignment. There was a very large increase in the amount of meaning-related comments they produced, with effect sizes that were consistently 1.5 standard deviation units or larger across the three key contrasts of this effect. There were still many more surface comments than meaning-related comments after training, but the ratio was much reduced from more than 100:1 before training to approximately 10:1 after training.

Finally, comparing the trained peers to the teacher at the direct contrast point of the third assignment, there was virtually no difference in the amount of surface comments produced, but the trained peers produced many more meaning-related comments. In general, across all the assignments, the exact amount of teacher-provided surface comments varied a moderate amount (means varying from 13 to 22), but the teacher consistently produced very few meaning-related comments. Similarly, the number of surface

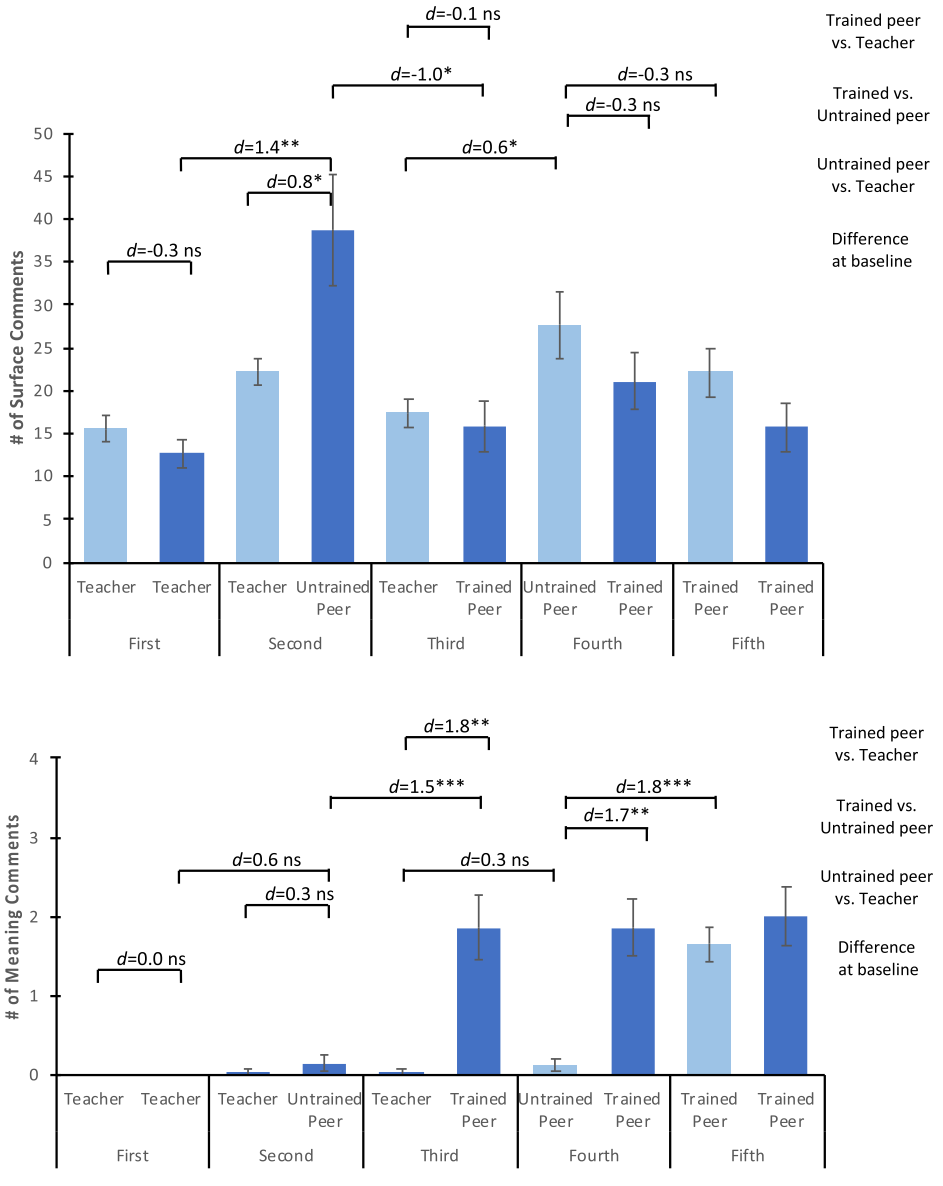

Figure 2. Mean (with SE bars) amounts of surface and meaning level feedback on the first draft in each condition (dark bar accelerated; light bar delayed) for each writing assignment, along with Cohen's *d* for the key contrasts indicated on the right of each graph. ***$p < .001$, **$p < 0.01$, *$p < 0.05$.

comments from untrained peers varied from assignment to assignment (means from 28 to 39), but very few meaning-related comments were produced.

The relationship between feedback and draft improvement

A series of multiple linear regressions were run to predict document 2 score, controlling for document 1 score (i.e., examine predictors of document improvements). To verify

Table 1. Means and standard deviations of the amounts of surface and meaning level feedback on the first draft in each condition in each writing assignment, along with the condition contrast Cohen's *d* and *t*-test value. $^{**}p < .01$, $^{*}p < .05$.

		Accelerated group		Delayed group			
Assignment	Level	*M*	SD	*M*	SD	*t*	Cohen's *d*
First	Surface	12.73	7.65	15.56	9.09	1.21	−0.34
	Meaning	0.00	0.00	0.00	0.00	–	–
Second	Surface	38.73	30.08	22.24	9.32	−2.50*	0.84
	Meaning	0.14	0.47	0.03	0.17	−1.03	0.34
Third	Surface	15.91	13.60	17.38	9.47	0.48	−0.13
	Meaning	1.86	1.88	0.03	0.17	−4.55**	1.79
Fourth	Surface	21.09	15.26	27.68	22.48	1.20	−0.35
	Meaning	1.86	1.64	0.12	0.41	−4.89**	1.70
Fifth	Surface	15.77	12.93	22.18	16.48	1.54	−0.45
	Meaning	2.00	1.77	1.65	1.30	−0.86	0.23

that feedback received was the main source of document improvement, an initial regression was conducted including the number of surface comments given, the number of meaning comments given, the number of surface comments received, and the number of meaning comments received, with assignment as a control variable. The η^2 values for the number of given comments variables were very small (<.01) and not statistically significant ($ps > .4$). Thus, for the remaining analyses, the number of comments provided were excluded from the regressions.

Next, a regression was conducted using data from all five assignments and combining all data sources. Both the number of surface and meaning comments received were significant predictors of document improvement, but the effect of surface comments received was much larger ($\eta^2 = .24$, $p < .001$) than the effect of meaning comments received ($\eta^2 = .02$, $p < .05$).

Then this regression was repeated separately for each feedback source (and always controlling for effects of writing assignment). As shown in Table 2, the number of meaning comments received was never a statistically significant predictor of document improvements, but the number of surface comments received was always statistically significant. It appeared that the teacher comments had the largest per-comment effect on revision improvements and that untrained peer comments had the smallest effect.

Table 2. Regression beta weights in predicting draft 2 scores using draft 1 scores and # of surface and meaning comments received (controlling for the effects of writing assignment). ns = not significant, $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

	Feedback source		
Predictor	Teacher	Untrained peer	Trained peer
Draft 1	0.97***	0.97***	0.92***
# of surface comments	0.03***	0.01***	0.02***
# of meaning comments	−0.11 ns	0.11 ns	0.03 ns
N	124	56	100
R^2	0.98	0.97	0.95

To more directly visualize the differential impact of each comment on revision quality, the number of received surface comments was binned into four categories to create four equally populated bins: Low (< 11 comments), Medium (between 11 and 18 comments), High (between 18 and 27 comments) and Very high (> 27 comments). A series of ANCOVAs predicting draft 2 score based on these four categories, controlling for draft 1 score and assignment, separately for each feedback source. The estimated draft 2 score was calculated for each bin category and graphed in Figure 3. Matching the regression results, the effect of surface comments was the weakest in the case of untrained peers. The overall effect of the amount of feedback was similar in the case of trained peers and teacher feedback, although the draft 2 score was higher for each amount of feedback level for trained peer feedback. This overall effect on the draft 2 score will be examined directly in the next section.

Effects of untrained/trained peer and teacher feedback on draft 2 quality

Figure 4 presents the estimated draft 2 quality after controlling for draft 1 quality in each condition on each assignment. Note that these estimated means are based on models comparing groups within an assignment. The mean estimates can be different for models comparing assignments within a group; the η^2 and statistical significance in each presented contrast are based on the relevant underlying model. Note also that similar findings were observed when simple different scores were used (draft 2 minus draft 1). As a baseline in the first assignment, the two classes showed equivalent draft 2 scores.

Looking across the three tests of the untrained peer vs. teacher contrast, the teacher feedback produced significantly better second drafts than did untrained peer feedback with small-to-large effect sizes. In the third case, the contrast was small and not statistically significant. These findings are generally consisted of the prior section analysis

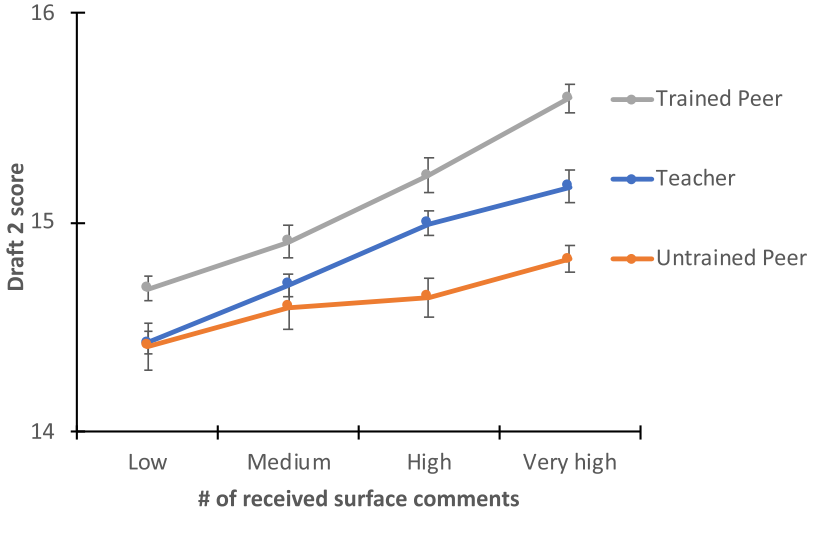

Figure 3. Mean estimated draft 2 score (with SE bars) as a function of different amounts of received surface comments in each feedback source, controlling for draft 1 score and assignment.

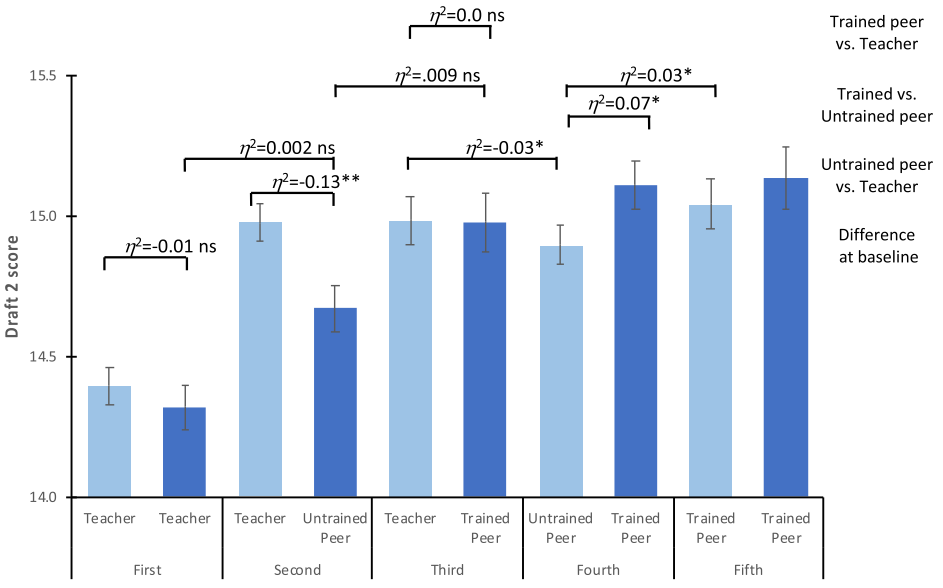

Figure 4. Estimated mean draft 2 scores (with SE bars) in each condition and each writing assignment, after controlling for draft 1 scores (dark bar accelerated; light bar delayed), along with η^2 for the key contrasts indicated on the right of each graph. ***$p < .001$, **$p < 0.01$, * $p < 0.05$.

showing a greater effect of teacher feedback than untrained peer feedback on draft 2 scores.

For two of the three tests of the untrained vs. trained peer contrast, feedback from trained peers produced significantly better revisions than did feedback from untrained peers, with small-to-medium effect sizes. The third contrast was not statistically significant. This contrast and the non-significant contrast in the teacher vs. untrained peers comparison were both within the accelerated group. This group had a smaller class size, and thus there may be a lack of power in these specific cases. Again, the overall pattern is similar to what was found in the prior section: feedback from trained peers appeared to lead to greater improvements that did feedback from untrained peers.

Finally, in the one teacher vs. trained peers contrast (for the third assignment), both feedback sources appeared to be exactly equivalent in terms of benefits for revision. This result also appears consistent with the findings of the previous section in which the benefit of each additional comment was roughly the same across conditions.

General discussion

This study sought to investigate the relative effectiveness of teacher feedback and both untrained and trained peer feedback in EFL writing classes. The study involved a training approach to improve student's peer feedback literacy that was modified to be detailed and responsive while also being scalable to teachers responsible for large numbers of students. The findings presented above confirmed results found in previous studies regarding the effectiveness of peer feedback training. First, peer feedback training produced much

more peer feedback at the meaning level, similar to what Berg (1999) found. This change is likely a result of both the training's discussion of meaning-level feedback and the guidance sheet given as part of the training. Although the writing assignments in the study were not designed to focus on meaning-level development, applying this approach in writing-to-learn contexts or with more complex texts should therefore produce learning benefits.

Secondly, peer feedback training led to greater improvement in terms of draft revisions, similarly to prior findings in both ESL and EFL writing contexts (Berg, 1999; Connor & Asenavage, 1994; Min, 2006). The effects of training on draft quality were sometime large, which is interesting given that much of the focus on the training was on meaning-level feedback, which was not a central issue in writing assignments included in this study. Thus, in addition to increasing the focus on meaning-level feedback, this training must also have improved other aspects of feedback quality.

Third, consistent with prior research on teacher vs. peer feedback (Hyland, 1998; Lee, 2007; Zhao, 2010), the study's teacher often gave vague feedback, which required learners to use the feedback without fully understanding it. That this finding is a replication of prior findings is important to address the weakness in the current study design: only one teacher was studied. While it is certainly possible to provide additional training to teachers that emphasizes the importance of detailed feedback, it is also important to recognize the logistical challenge of requiring detailed teacher feedback in the context of high teaching workloads.

In addition to replicating prior findings on feedback training, the current study also added additional information. For example, the current study found that peer feedback after training had a similar effect as teacher feedback on draft improvement. Such a finding is significant because prior working comparing peer to teacher feedback focused on untrained peer feedback, which provided an overly negative assessment of the potential of peer feedback (Connor & Asenavage, 1994; Paulus, 1999; Yang et al., 2006). The current study shows that training is necessary for peer feedback practice in EFL contexts to be able to obtain benefits that are equivalent to teacher feedback.

Another interesting observation in the current study is that peer reviewers produced more feedback at the surface level before training than did the teacher. One likely factor is that repeated feedback was not eliminated in the analyses. However, repeated feedback is not necessarily wasteful, as some recent research suggests that students are more likely to make revisions when they receive multiple comments on a particular problem (Gao et al., 2018; Wu & Schunn, 2020).

A third interesting finding is that the number of surface-level comments produced by the teacher, untrained peer reviewers, and trained peer reviewers all were positively associated with improvements in revision quality, whereas the number of meaning-level comments had no significant relationship with revision quality. Berg (1999) found that peer feedback training led to more meaning-type revisions, which were important for draft improvements. It is important to realize that the participants in the current study had relatively low English proficiency. Therefore, there were many language problems in their documents. Indeed, interviews with students suggested that especially the lower English-proficiency learners preferred feedback on surface-level issues. It may be that the motivations that students have regarding learning in their

EFL or ESL classrooms may shape both what they produce and how they respond to received feedback.

Caveats and future research

While the current study had the advantage of combining within and between-group contrasts, there are important experimental limitations that should be acknowledged. First, the assignment of participants to condition was at the class level and there were only two classes. Thus, inherent differences between classes could have contributed to some of the between-group effects. However, the two conditions were equivalent in writing performance at the baseline first assignment and the condition contrast findings generally replicated across within-group and between-group contrasts.

Second, the study simultaneously manipulated the receiving and providing aspects of peer feedback, which can limit inference about which component contributed to greater draft revision. However, such a combined change approach is more naturalistic, and regression analyses suggest that relatively little of the draft changes came from the providing feedback component. Future studies using assignments that are more focused on meaning-level development likely will find benefits from the providing aspects (e.g., Lundstrom & Baker, 2009).

Third, the study was conducted in one specific context: 3rd year English majors at a particular university. Future research will have to be conducted to see how the training effects replicate across contexts, particularly with students who have a stronger mastery of English. Similarly, the training study should be replicated with non-English majors who might have lower interest/motivation in learning English, which can affect attitudes towards participation in peer feedback (Zou et al., 2018).

Fourth, on a related aspect, the current study involved learning-to-write assignments of a certain general form. These kinds of writing assignments placed heavy emphasis on the foundations of written English. The training approach which placed heavy emphasis on meaning-level feedback might have had even larger effects on writing-to-learn assignments or more advanced forms of writing (e.g., a research proposal).

Conclusion

We draw three main conclusions from this study. Firstly, peer reviewers, after training, can provide more meaning-level feedback than do teachers in high workload contexts. Secondly, peer feedback with training can have a similar effect on revision as teacher feedback does (in high workload contexts). Thirdly, students, after training, produce feedback that leads to more revision; a key result for any context using peer feedback for pedagogical or workload reasons.

The current study demonstrates that it is feasible to provide effective peer feedback training in the context of EFL students who are relatively weak in English and under typical workload conditions. Therefore, peer feedback, after training, is a suitable pedagogy for foreign language teaching because it can help to relieve the heavy workload experienced by those teachers and produces effective feedback. Moreover, it can be easily implemented using the many web-based peer feedback systems. Finally, our

findings suggest that peer feedback training is critical to observing the benefits of peer feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Social Science Foundation of China (BJA200097).

ORCID

Ying Cui http://orcid.org/0000-0002-2668-2092
Christian D. Schunn http://orcid.org/0000-0003-3589-297X
Xiaosong Gai http://orcid.org/0000-0002-5178-8796

References

Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research & Development*, *37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913

Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research*, *20*(4), 498–513. https://doi.org/10.1177/1362168814561902

Asparo, C. A., Mercader, C., & Ion, G. (in press). Making peer-feedback more efficient: What conditions of its delivery make the difference? *Higher Education Research & Development*. https://doi.org/10.1080/07294360.2020.1840527

Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, *8*(3), 215–241. https://doi.org/10.1016/S1060-3743(99)80115-5

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carson, J. G., & Nelson, G. L. (1994). Writing groups: Cross-cultural issues. *Journal of Second Language Writing*, *3*(1), 17–30. https://doi.org/10.1016/1060-3743(94)90003-5

Cho, K., & Schunn, C. D. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891

Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, *3*(3), 257–276. https://doi.org/10.1016/1060-3743(94)90019-1

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*(4), 400–414. https://doi.org/10.2307/356602

Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge University Press.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*(3), 161–184. https://doi.org/10.1016/S1060-3743(01)00039-X

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, *31*(2), 315–319. https://doi.org/10.2307/3588049

Gao, Y., Schunn, C. D., & Yu, Q. (2018). The alignment of written peer feedback with draft problems and its impact on revision in peer assessment. *Assessment and Evaluation in Higher Education*, *44*(2), 294–308. https://doi.org/10.1080/02602938.2018.1499075

Ghani, M., & Asgher, T. (2012). Effects of teacher and peer feedback on students' writing at secondary level. *Journal of Educational Research*, *15*(2), 84–97.

Guerrero, M., & Villamil, O. (1994). Socio-cognitive dimensions of interaction in L2 peer revision. *Modern Language Journal*, *78*(4), 484–496. https://doi.org/10.1111/j.1540-4781.1994.tb02065.x

Ho, P. V. P., Phung, L. T. K., Oanh, T. T. T., & Giao, N. Q. (2020). Should peer E-comments replace traditional peer comments? *International Journal of Instruction*, *13*(1), 295–314. https://doi.org/10.29333/iji.2020.13120a

Huisman, B., Saab, N., Driel, J. V., & Broek, P. (2017). Peer feedback on college students' writing: Exploring the relation between students' ability match, feedback quality and essay performance. *Higher Education Research & Development*, *36*(7), 1433–1447. https://doi.org/10.1080/07294360.2017.1325854

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: A meta-analysis. *Assessment & Evaluation in Higher Education*, *44*(6), 863–880. https://doi.org/10.1080/02602938.2018.1545896

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing*, *7*(3), 255–286. https://doi.org/10.1016/S1060-3743(98)90017-0

Kroll, B., & Vann, R. (1981). *Exploring speaking-writing relationships: Connections and contrasts*. National Council of Teachers of English.

Lam, R. (2010). A peer review training workshop: Coaching students to give and evaluate peer feedback. *TESL Canada Journal*, *27*(2), 114–127. https://doi.org/10.18806/tesl.v27i2.1052

Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing*, *12*(3), 180–198. https://doi.org/10.1016/j.asw.2008.02.003

Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57–68). Cambridge University Press.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Ma, J. (2018). Usability of teacher written feedback: Exploring feedback practices and perceptions of teachers and students. *Electronic Journal of Foreign Language Teaching*, *15*(1), 23–38.

Min, H. T. (2005). Training students to become successful peer reviewers. *System*, *33*(2), 293–308. https://doi.org/10.1016/j.system.2004.11.003

Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118–141. https://doi.org/10.1016/j.jslw.2006.01.003

Nelson, G., & Murphy, J. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, *27*(1), 135–142. https://doi.org/10.2307/3586965

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, *8*(3), 265–289. https://doi.org/10.1016/S1060-3743(99)80117-9

Ren, H., & Hu, G. (2012). Peer review and Chinese EFL/ESL student writers. *English Australia Journal*, *27*(2), 3–16.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, *59*(1), 23–30. https://doi.org/10.1093/elt/cci003

Ruegg, R. (2015). The relative effects of peer and teacher feedback on improvement in EFL students' writing ability. *Linguistics and Education*, *29*, 73–82. https://doi.org/10.1016/j.linged.2014.12.001

Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, *8*(1), 31–54. https://doi.org/10.1191/1362168804lr133oa

Schillings, M., Roebertsen, H., Savelberg, H., Dijk, A. V., & Dolmans, D. (2020). Improving the understanding of written peer feedback through face-to-face peer dialogue: Students' perspective. *Higher Education Research & Development*, *40*(5), 1100–1116. https://doi.org/10.1080/07294360.2020.1798889

Schunn, C., Godley, A., & DeMartino, S. (2016). The reliability and validity of peer review of writing in high school AP English classes. *Journal of Adolescent & Adult Literacy*, *60*(1), 2–11. https://doi.org/10.1002/jaal.525

Snowball, J. D., & Mostert, M. (2013). Dancing with the devil: Formative peer assessment and academic performance. *Higher Education Research & Development*, *32*(4), 646–659. https://doi.org/10.1080/07294360.2012.705262

Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, *1*(3), 217–233. https://doi.org/10.1016/1060-3743(92)90004-9

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, *9*(2), 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9

Villamil, O., & Guerrero, M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, *19*(4), 491–514. https://doi.org/10.1093/applin/19.4.491

Wu, Y., & Schunn, C. D. (2020). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology*, *62*, 101897. https://doi.org/10.1016/j.cedpsych.2020.101897

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*(3), 179–200. https://doi.org/10.1016/j.jslw.2006.09.004

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, *4*(4), 209–222. https://doi.org/10.1016/1060-3743(95)90010-1

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, *15*(1), 3–17. https://doi.org/10.1016/j.asw.2010.01.002

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development*, *37*(4), 883–897. https://doi.org/10.1080/07294360.2018.1446417

Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication*, *12*(4), 492–528. https://doi.org/10.1177/0741088395012004004

Zou, M., Schunn, C., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment and Evaluation in Higher Education*, *43*(5), 800–811. https://doi.org/10.1080/02602938.2017.1409872

Appendix. Peer review sheet

Please answer the following questions with the purpose to provide your classmate with honest but helpful reactions and responses as the reader of this essay.

(1) What do you like the best about the ideas in this essay? Be specific. (Precise vocabulary, cohesive/linked ideas, clear/easy to follow, convincing, effective reasoning, well-developed ideas, attention-grabbing introduction, strong conclusion, intriguing style, well-supported topic sentences, understandable transitions, etc.)

(2) Can you find a thesis statement? Can you find a clear topic sentence in each body paragraph? If you can, in your own words, state the focus/thesis/topic of the writing.

(3) How many reasons and supporting proof are provided? Do all of these reasons logically support the writer's opinion? Explain how well do these reasons persuade you that the author's opinion is the correct one?
(4) Are there any ideas in the essay that are not clear or that you find confusing? Mark these with a letter C. Explain why you think this is confusing and make some suggestions for improvement.
(5) Are there any ideas in the essay that need further development? About which parts of the essay would you like more information? Mark these with a letter D. Explain why you think this should be developed more and make some suggestions.
(6) How effective is the conclusion?
(7) What questions, comments, and/or suggestions do you have for the author?