

The alignment of written peer feedback with draft problems and its impact on revision in peer assessment

Ying Gao, Christian Dieter D. Schunn & Qiuying Yu

To cite this article: Ying Gao, Christian Dieter D. Schunn & Qiuying Yu (2019) The alignment of written peer feedback with draft problems and its impact on revision in peer assessment, *Assessment & Evaluation in Higher Education*, 44:2, 294-308, DOI: [10.1080/02602938.2018.1499075](https://doi.org/10.1080/02602938.2018.1499075)

To link to this article: <https://doi.org/10.1080/02602938.2018.1499075>



Published online: 26 Sep 2018.



Submit your article to this journal [↗](#)





Article views: 91



View Crossmark data [↗](#)



The alignment of written peer feedback with draft problems and its impact on revision in peer assessment

Ying Gao^{a,b} , Christian Dieter D. Schunn^b  and Qiuying Yu^{a,b}

^aSchool of Foreign Languages, Northeast Normal University, Changchun, China; ^bLearning Research and Development Center (LRDC), University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

Despite the wide use of peer assessment, questions about the helpfulness of peer feedback are frequently raised. In particular, it is unknown whether, how and to what extent peer feedback can help solve problems in initial texts in complex writing tasks. We investigated this research gap by focusing on the case of writing literature reviews in an academic writing course. The dataset includes two drafts from 21 students, sampled to represent a wide range of document qualities, and 84 anonymous peer reviews, involving 1,289 idea units. Our study revealed that: (1) at both substance and high prose levels, drafts of all quality levels demonstrated more common problems on advanced writing issues (e.g. *counter-argument*); (2) peer feedback was driven by difficulty of the problem rather than overall draft quality, peer comments were not well aligned with the relative frequency of problems, more comments were given to less difficult problems; (3) peer feedback had a moderate impact on revision, and importantly, receiving multiple comments on the same issue led to more repairs and improvement of draft quality, but consistent with the comments received, authors tended to fix basic problems more often. Implications for practice and research are drawn from these findings.

KEYWORDS

Problems in writing; peer comments; repair rates; literature review writing

Introduction

Peer assessment/review/feedback has been recommended for more than four decades (Bruffee 1980; Chang 2016), and the last three decades have witnessed a growing body of research in contexts of both first (L1) and second language (L2) writing (Hyland and Hyland 2006; Min 2008; Hu and Lam 2010; Zheng 2012; Yu and Lee 2016). However, peer feedback is still a 'contentious and confusing issue throughout higher education institutions' (Boud and Molloy 2013: 698). In particular, questions about the helpfulness of peer feedback are frequently raised. Within L2 studies, research on peer review in the last two decades has covered various issues relating to the perceptions, process and product of peer review (Li et al. 2016; Zou et al. 2018). At the process and product ends of research, previous studies have in general followed a writer/reviewer/comment-centric perspective, focusing on the features of comments as a whole and the effects of the reviewing process (Tsui and Ng 2000; Lundstrom and Baker 2009; Cho and Cho 2011; Zhao 2014; Berggren 2015; Walker 2015; Huisman et al. 2018). Few studies have been conducted from a text-centric aspect which begins with an analysis of the problems in an author's text.

Therefore, the extent to which students can address and help solve problems in initial texts is largely unknown.

This issue is especially relevant in longer and complex writing tasks, where refining critical flaws in the central argument is more important than fixing peripheral aspects of the document. Unfortunately, most peer feedback studies in L2 have been conducted on short writing tasks (e.g. 100–300 words long) where the focus was on language errors or minor content issues (e.g. Yang, Badger, and Yu 2006; Allen and Mills 2016; Yu and Hu 2016; Chong 2017). Longer writing involving difficult content, like writing literature reviews (Flowerdew 2000; Kwan 2006), has received little attention. However, such writing is critical in academic learning, and it is particularly challenging for L2 writers as they have to struggle with an extra load related to language issues. As L2 peer feedback has been found to focus more on global features of writing (Tuzi 2004; Min 2005; Yang, Badger, and Yu 2006) and provide global comments (the development of ideas, audience and purpose, and organisation of writing; Chen 2010), it is potentially a helpful instructional method in such writing tasks.

Therefore, in this quasi-experimental design study of Chinese students writing literature reviews in English, we evaluate the features and problems in initial drafts, the alignment of peer comments with the problems, and the subsequent repair rate in second drafts in relation to the targeted comments received. Specifically, we focus on the substance of arguments and high prose issues, which are particularly challenging in this kind of writing.

Peer feedback as an important facilitator in writing

The use of peer feedback in L2 writing instruction has been supported by a number of theories like process writing theory, collaborative learning theory, interactionist theory in second language acquisition and cognitive and psycholinguistic fields, and sociocultural theory (Yu and Lee 2016). For example, it is believed that peer feedback can help achieve the transition from inter-psychological to intrapsychological functioning and help student writers move from stages of other-regulation to self-regulation (Villamil and De Guerrero 2006).

In academic writing, as in any other kinds of writing, writers in general have difficulty perceiving errors in their own writing compared to others' texts because errors in one's own text are often automatically mentally corrected (Flower et al. 1986). In writing literature reviews, novice writers often fail to include strong support for their hypothesis, include obvious or unsupported arguments, or fail to discuss counter-evidence (Schwarz et al. 2003; Nussbaum and Schraw 2007) because they do not know that a literature review serves as an argument for a hypothesis rather than just a historical summary (Barstow et al. 2017). ESL/EFL (English as a second/foreign language) learners, in particular, have difficulty in establishing a critical stance in literature reviews, and prefer an indirect and less critical way to express evaluations (Hinkel 1997). It is here where peer assessment could be applied since it can be as effective as an instructor's feedback in helping students improve their draft (Liu and Hansen 2002; Cho and Schunn 2007; Chen 2010). Multi-peer assessment can provide more total feedback than from an over-taxed instructor, more persuasive feedback when multiple reviewers note the same problems, and feedback representing more diverse audience perspectives (Liu and Hansen 2002; Tuzi 2004).

According to an Identical Elements Theory framing of providing feedback, the process and features of providing feedback are determined by both text quality and reviewer ability (Patchan and Schunn 2015). For example, higher-quality texts receive more praise and fewer criticisms (Cho and Cho 2011). However, although it is generally assumed that since low quality texts have more problems than high-quality texts and thus provide more opportunities for problem detection, diagnosis and selection of appropriate solutions, no significant effects of text quality were found in Patchan and Schunn's (2015) study. They concluded that relative quality, more than absolute quality, seemed to drive comment content. Therefore, despite the apparent impact of

text quality on peer feedback, no previous studies have documented the relationship between specific text qualities and peer feedback. For example, are there systematic patterns in which kinds of problems tend to receive feedback and which kinds of problems tend to be overlooked? Clearly, the focus of feedback will affect what knowledge is reinforced or gained and provide instruction for the writer (Lundstrom and Baker 2009; Ruegg 2015).

In addition, academic writing is not just a matter of addressing general high-level issues like organisation and genre-following. But also it involves attention to features specific to the genre (e.g. move structures in literature reviews, Kwan 2006). It is unclear whether peers, especially EFL writers, can consistently identify issues and suggest improvements in the complex aspects of academic writing. The problems found in weaker students' texts might be more readily discovered, whereas the problems found in stronger students' writing may be beyond the peers to diagnose and repair. It is therefore important to find out the concerns of the reviewers in relation to the text being reviewed, an analytic approach rarely used in research on peer assessment.

The current study

The goal of the current study was to understand the relative impact of peer assessment for EFL academic writing, focusing on the case of literature review writing. Although author ability is a determining factor, text quality is influenced by a number of factors such as topical interest, grade goals set for the course, and competing course and work demands (Schiefele 1999). Thus, author ability may not have as large an effect on the quality of documents being reviewed as one might have otherwise expected (Patchan and Schunn 2016). Therefore, this study examines text quality rather than author level per se. The study intends to answer three research questions, one specific to literature review writing, and two more generally about peer review of academic writing:

1. What are common problems in literature reviews written by EFL writers, specifically in substance (move structures) and general high prose levels?
2. Do EFL peer feedback comments tend to focus on problems at a particular difficulty level?
3. Does the relevant emphasis within peer feedback explain which kinds of problems tend to be addressed in revisions?

Research context

The study was conducted within a senior year course entitled 'Academic Writing' in a BA programme in English at a large public research university in China. The course teaches academic writing conventions and prepares students for writing a graduation thesis. Students were trained to use a peer review system *Peerceptiv* (based on *SWoRD*) (Schunn 2016). Within the system, students' first and second drafts were each randomly distributed to four peers for rating and comments. Prior to the peer review assignment, students had been informed about general conventions in academic writing, and they had been given instruction on writing literature reviews.

Participants

Participants in the course were native speakers of Mandarin Chinese and had learned English for at least 12 years on average. Purposive sampling based on four criteria was used in selecting the participants from a total of 126 students taking the course. (1) Fulfilment of assignment. Nineteen students who did not submit either both drafts or submitted substantially plagiarised documents (identified using a tool) were removed. (2) Common research area. To allow for a

detailed domain-specific scheme for analysis, analyses focused on the 60 papers with a shared general topic of linguistics, dropping the 47 papers on topics of literature or translation. (3) Diversity of writing quality. A representative set of 10 papers were selected from each of the high (77%–87%), mid (66%–71%) and low (44%–60%) groups by taking only the odd-numbered papers from the list of papers that was ordered by score. (4) Number of comments received. Focusing on papers that received comments from four reviewers resulted in seven documents for each group (papers receiving comments from fewer than four reviewers were deleted to remove a source of variance on paper revision behaviour that was not the focus of the current study). So finally, the dataset included 21 first drafts (seven from high, mid and low group, respectively), each with comments from four peer reviewers, and 21 second drafts.

Fifty-eight students (20–22 years old) were authors and/or reviewers for this dataset, and they all agreed that their data could be used for research. Their English proficiency was approximately between 75 and 105 on the Test of English as a Foreign Language examination. They had received training for three years on English writing before taking this academic writing course.

Procedure

Participants completed three main tasks in this study: (1) they wrote a first draft, (2) they reviewed peers' texts and (3) they revised own text based on peer feedback. In the assignment, students were asked to write a literature review (word limit: 1,000 words) of a study they expected to complete as their senior thesis. A research gap or an unanswered question in the literature was supposed to be described, and a critique that assessed the value of relevant theories, ideas, claims, research designs, methods or conclusions of the selected papers was expected. Students submitted two drafts in weeks 15 (draft 1) and 19 (draft 2) of the term. In the four weeks between draft submissions, they reviewed four peers' papers, revised their own documents after receiving peer feedback, and received weekly instruction by the course teacher through demonstrating good commenting and revising. To guarantee peer comments which were based solely on the text itself rather than personal relationships or prior experience with text drafts, as well as writers' incorporation of peer feedback that was determined by the feedback quality itself (Cote 2014), documents were randomly distributed among peers and both authors' and reviewers' identities were blinded.

Reviewing rubric

To support the review process, students were given a detailed reviewing rubric related to four dimensions of the documents: introduction, main body, conclusion, and rules and conventions. In each of these four dimensions, reviewers were required to give at least one comment and one rating on a 7-point Likert scale running from 'difficult to read, poor, fair, OK, good, very good, to excellent'. The Likert scales had short descriptions for each rating level (e.g. 'excellent – presents the literature in a highly critical and objective manner, with clear focus and problem involvement by comparison and contrast'). Following Patchan et al. (2013), written feedback was in the form of end comments (i.e. comments separated from the text) rather than marginalia (i.e. comments located in the margins of the text).

Measures

Text quality

Text quality was based on multiple peer ratings (four each), which has proven to be a reliable means of document evaluation, sometimes even more reliable than a single tutor's ratings (Liu and Carless 2006; Cho and Schunn 2007). Peer ratings were conducted from the equally

Table 1. Coding schemes for literature review drafts and peer comments.

Dimensions	Features	
Substance (thematic moves)	Move 1: Establishing the territory of one's research by:	
	Topic and argument	Presenting central topic, clarifying argument
	Survey of research	Surveying the research-related phenomena
	Necessity of investigation	Necessity/significance of further investigation
	Move 2: Creating a research niche by:	
	Counter-argument	Presenting counter-claiming/argument
	Critical evaluation	Criticizing/evaluating knowledge or research surveyed
	Own research	Relating the surveyed claims to one's own research
	Move 3: Occupying the research niche by announcing:	
	Gap identification	Identification of research gap
High prose (text quality)	Aim of research	Research aims, research questions or hypotheses
	Coverage of issues	Coverage of issues in central topic
	Clarity of writing	Clear presentation of thesis or evidence
	Transition and organization	Textual coherence and use of discourse markers like 'so, because, therefore ...'
	Effectiveness of argument	Strength of argument in terms of correctness, sufficiency, depth, consistency

weighted dimensions on 7-point Likert scales, and total score was converted to a percentage. To establish the validity of the mean peer ratings, two writing experts (i.e. graduate students in applied linguistics with writing teaching experience) rated the quality of the students' first and second draft using a 100-point scale based on the reviewing rubric provided to the students. Inter-rater reliability between the two expert raters was high ($r=0.79$), as was the correlation between peer ratings and expert ratings ($r=0.79$), all $ps < 0.05$. The one-way analysis of variance (ANOVA) test on peer ratings of the three groups of drafts confirmed large significant differences in draft quality: high-rating writings ($M=82.3$, $SD=2.2$), mid-rating writings ($M=68.9$, $SD=1.3$) and low-rating writings ($M=53.7$, $SD=5.0$), $F(2,20)=137.6$, $p < 0.00001$.

Text features

To unpack the details of feedback provided and revisions made in the documents, coordinated coding schemes (see Table 1) were developed that characterised both documents and comments in terms of the substance of literature reviews, and two dimensions applicable to all academic writing (high prose and low prose). The substance dimension was based on the three-move structure proposed by Kwan (2006), with adaptations made on the basis of the reviewing rubric given by the course teachers. Kwan's three-move structure was followed since it was based on an adapted CARS (create a research space) model (Swales 1990, 2004), and her analysis of the rhetorical structure of literature review chapters in PhD theses in applied linguistics, which is very similar to the context and genre of data for the present study. High prose refers to high-level writing issues which involve clarity, organisation of writing, and effectiveness or strength of argument. Low prose refers to an issue dealing with the literal text choice – usually at a lexical or a syntactic level. However, although all authors had writing problems at the low prose level, peer comments at this level were rather scarce on these issues and not repaired in most cases. Considering our focus on substance and high prose issues, which are more likely to improve the quality of the text when implemented (Lundstrom and Baker 2009; Rahimi 2013; Patchan and Schunn 2016), comments on low prose issues were eliminated from further analysis.

Draft coding

Coding of problems in draft 1 and draft 2 was conducted by two expert EFL teachers. Documents were coded for each substance and high prose issue in binary terms of having the

Table 2. Mean and standard deviations for document length and quality for each draft, overall and by draft quality group.

N		Mean		SD		
		Draft 1	Draft 2	Draft 1	Draft 2	
	Draft length (word count)	21	1,222.3	1,585.6	260.1	529.0
	Draft quality (rating in %)	21	68.0	74.9	12.3	8.7
	High quality	7	82.3	83.6	2.2	4.4
	Mid quality	7	68.9	75.7	1.3	4.4
	Low quality	7	53.7	65.4	5.0	4.7

specific problem type or not. *Kappa* for inter-rater reliability was high in both drafts ($K=0.81$ and $K=0.83$).

Feedback coding

Feedback comments were segmented by idea unit because reviewers sometimes discussed more than one idea within a single comment entry. Following Nelson and Schunn (2009), an idea unit was defined as a contiguous comment referring to a single topic, the length of which can vary from a few words to several sentences. Comments were coded by two independent coders (the same two experts who coded the drafts). *Kappa* values for each of the 12 coding categories ranged from 0.66 to 1, with most of them being above 0.8, indicating high inter-rater reliability. To minimise coding noise, all documents and comments were double-coded, and conflicts were resolved through discussion.

Data description and analysis

In total, we examined 21 first drafts and 21 second drafts (see Table 2). The comments were on draft 1 only. Each document was reviewed by four reviewers and each review could consist of multiple idea units along each of four dimensions; collectively, the comment dataset involved 1,289 idea units. Second drafts were both longer and of higher quality, providing a meaningful context for studying substantive document change due to peer feedback.

To answer our three research questions, three steps were taken. First, problem frequencies in initial drafts were identified across categories and groups. Second, comment category frequencies were measured, and correlations by category and ANOVA by quality group were run to see how well comments aligned with actual problems. Third, repair rates by categories were determined and the impact of comment frequency on revision was assessed with *t*-tests and Cohen's *d*. As a text-centric and specifically problem-centric study, all analysis of data began from and centred around writing problems in authors' initial drafts.

Results

Results of the study are presented by following the order of our three research questions.

Identifying initial challenges (problems) in writing (RQ1)

A majority (52%) of the 21 drafts had problems at both substance and high prose levels, and the ability groups varied substantially both in whether they had both kinds of problems and the number of problems: low (82%, $M=5.3$, $SD=2.0$), mid (51%, $M=3.3$, $SD=2.5$), and high (23%, $M=1.5$, $SD=1.8$), $F(2,35)=9.5$, $p < 0.001$.

As seen in Figure 1, the low ability group commonly had problems in every category. A majority of the mid-quality group suffered problems in only three categories, with lower frequencies in other categories, including no problems in the *survey of research* category. A majority of

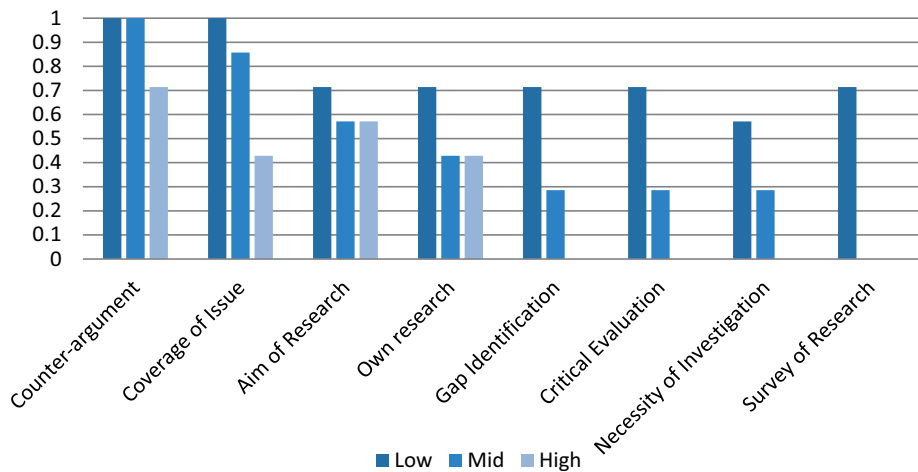


Figure 1. Draft 1 problem rate for substance issues across draft quality groups.

the high-quality group had problems in only two categories, and no problems at all for four categories.

But there were also consistent patterns across groups. Among the three moves, move 1 appeared least challenging to all authors as most authors had fulfilled the task of introducing the *topic & argument* (no problem occurred and therefore deleted from analysis), reviewing previous research, and summarising the necessity or importance of research on the topic. By contrast, 59% and 57% of all authors had problems with moves 2 and 3. In particular, the largest proportion of authors had problems with *counter-argument* (90%), *coverage of issues* (76%) and *aim of research* (62%). Authors appeared to be avoiding any mention of these aspects of literature reviews (advanced writing issues), which are generally more demanding in that the author must produce their own interpretation and evaluation of the literature, along with a description and justification of their own research.

Figure 2 indicates that 94% of the low group and 74% of the mid group had problems with all the three categories; 22% of the high group had problems with two categories, but not with *clarity of writing*. However, the writings of all groups had problems in *effectiveness of argument* and *transition and organisation*.

In sum, text quality groups varied substantially in total number and in particular kinds of problems in the literature review writings. Particular kinds of problems varied substantially in base-rates, ranging from those found in all literature reviews to those found only in a subset of low quality literature reviews.

The alignment of peer comments with problems in initial drafts (RQ2)

In general, peer feedback addressed all of the topic areas, but in varying frequencies per document that was not always well aligned with the relative frequency of problems. Across all comments received by a text, some issues were addressed less than once per document on average, and others were addressed many times.

Failure of comments in addressing problems

Overall, high prose problems ($M = 4.9$, $SD = 4.4$) received more comments than substance problems ($M = 2.3$, $SD = 2.7$), $t(46.8) = 3.4$, $p < 0.001$, Cohen's $d = .71$. Within substance, the most comments were focused upon *own research* ($M = 4.4$), and the fewest were given to *aim of research* ($M = 0.5$) and *counter-argument* ($M = 0.6$), $F(2,42) = 14.9$, $p < 0.0001$. In terms of high prose issues,

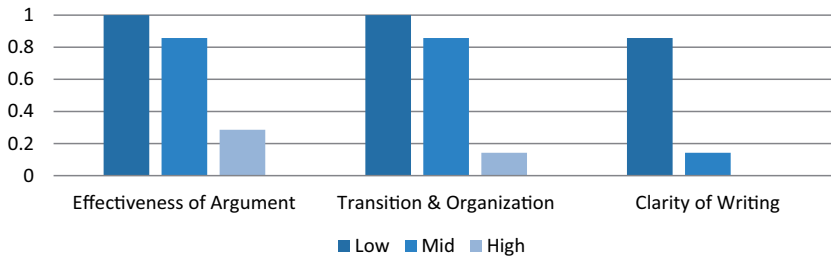


Figure 2. Draft 1 problem rate for high prose issues across draft quality groups.

many comments were given to *transition and organisation* ($M = 7.6$, $SD = 5.0$), and relatively fewer comments were given to *effectiveness of argument* ($M = 2.7$, $SD = 3.0$), $t(27) = 3.3$, $p < 0.01$, $d = 1.19$.

Unfortunately, the most challenging areas received the least comments. That is to say, writings received more comments on issues with fewer problems and fewer comments on issues with more problems (see Figure 3). *Counter-argument*, in which 79% of the literature reviews had problems, received a very low mean of 0.6 comments ($SD = 0.9$), whereas *survey of research*, in which only 21% of the literature reviews had problems, received a high mean of 4.0 comments ($SD = 2.4$), $t(4.3) = 3.2$, $p < 0.05$, $d = 1.9$. The only exception was *transition and organization* (58% problem rate, 7.6 mean comments). Thus, reviewers were more likely to attend to issues that were not difficult for them as writers. Correlation analysis at the document level (i.e. having problems or not with the number of comments received) showed that only in four categories out of 11 were they positively related, and for all of them r was below 0.2, all with $ps > 0.2$. That is, there was little relationship between having a problem and the number of comments received. But for issues that were problematic for less than half of the documents, authors were provided with a mean of at least three relevant comments from peers.

In both writing and reviewing, the 'basic' elements in writing literature reviews (like *survey of research*, *necessity of investigation*, *clarity of writing*) were given more attention while the 'advanced' elements (like *counter-argument*, *aim of research*, *coverage of issue* and *effectiveness of argument*) were detected less often. The latter, of course, is more demanding on writer/reviewer's knowledge and writing strategies, a potential factor explaining why novice writers are very often found only summarising and presenting part of the story in the challenging task of reviewing the literature (Hinkel 1997; Nussbaum and Schraw 2007; Barstow et al. 2017).

Comments for strong and weak texts

Similar to prior research in that lower-quality texts received more criticism comments than higher-quality texts (Patchan, Charney, and Schunn 2009; Patchan et al. 2013; Patchan and Schunn 2016), in this study, weaker writings received directionally more comments: low ($M = 3.4$), mid ($M = 3.2$) and high ($M = 1.8$), but the difference was small and not statistically significant, $F(2,119) = 1.52$, $p > 0.1$. Thus, it is likely the difficulty of the problem that tended to receive comments and drove the amount of feedback, rather than overall document quality per se.

Helpfulness of peer feedback in revision (RQ3)

A paired t-test between draft 1 ($M = 68.0$, $SD = 12.3$) and draft 2 ($M = 74.9$, $SD = 8.7$) ratings indicated significant improvement in draft quality after peer feedback and revision, $t(20) = 4.2$, $p < 0.001$, with a medium effect size ($d = .65$). Students generally repaired 1/4 (24%) of the issues,

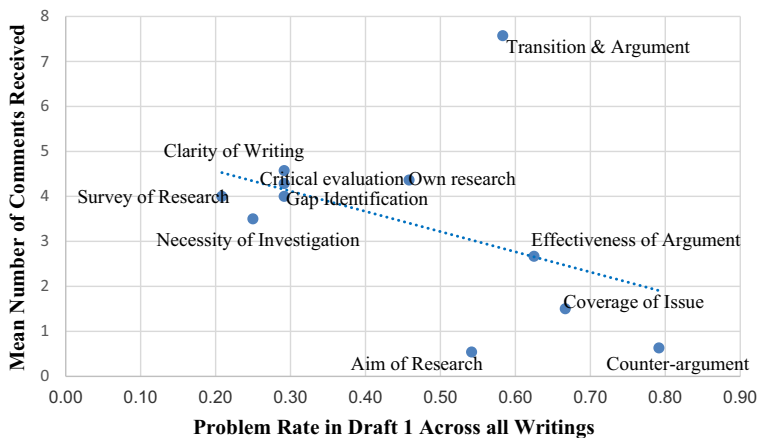


Figure 3. Relative frequency for each problem in author drafts and comments received from peers.

with roughly similar success rates across quality groups (between 21% for high and 26% for low).

Identifying success and failure in revision with peer feedback

However, issues varied widely in terms of how likely they were to be addressed in revision, from 0% (*gap identification*) to 100% (*survey of research*). In general, substance issues were repaired more often (36%) than were high prose issues (21%), but not significantly so, $t(118) = 0.8$, $p > 0.1$. In general, students were more likely to repair the less challenging problems (see Figure 4). Within substance, 92% of problems in move 1 (the easiest) were fixed, whereas only 23% and 12% of problems in move 2 and move 3 were fixed, $F(2,83) = 19.9$, $p < 0.00001$. The highest rate of repair was on *survey of literature* (100%) and *necessity of investigation* (83%), where only 21% and 25% of problems existed, whereas in cases where most writings had problems, like *counter-argument* (79%) and *coverage of issue* (67%), only 5% and 6% were repaired, $F(1,45) = 95.6$, $p < 0.00001$. Over all, repair rate and problem rate were negatively correlated ($r = -0.63$, $p < 0.05$).

Correlation between comments and repair

On a by-issue basis, repair rate and comment frequency were positively but weakly correlated ($r = 0.33$, $p < 0.05$). As shown in Figure 4, having a high commenting rate for the issue was necessary but not sufficient to produce a high repair rate. For example, issues noted by fewer than 80% of the peers were never repaired by a majority of students. However, only two (of the five) issues mentioned by more than 80% of the reviewers were commonly repaired.

Amount of feedback appeared to also matter: authors who fixed their problems received 4.3 comments ($SD = 3.7$), whereas authors who did not fix their problems received 2.7 comments ($SD = 3.4$), $t(118) = 2.2$, $p < 0.05$ ($d = .46$), a medium-sized effect. This suggests that, in general, authors were likely to neglect peer comments, but when they received relatively more comments, they were more likely to fix the problems. This effect was large for both the mid ($d = 1.3$) and high groups ($d = 1.1$), and almost zero for the low group ($d = 0.05$), suggesting that mid and high group authors were more influenced by the number of comments received, whereas low group authors were not.

A more careful observation of this relationship revealed that peers benefited from having multiple comments on an issue, $\chi^2(3) = 8.90$, $p < 0.05$ (see Figure 5). Having only one comment

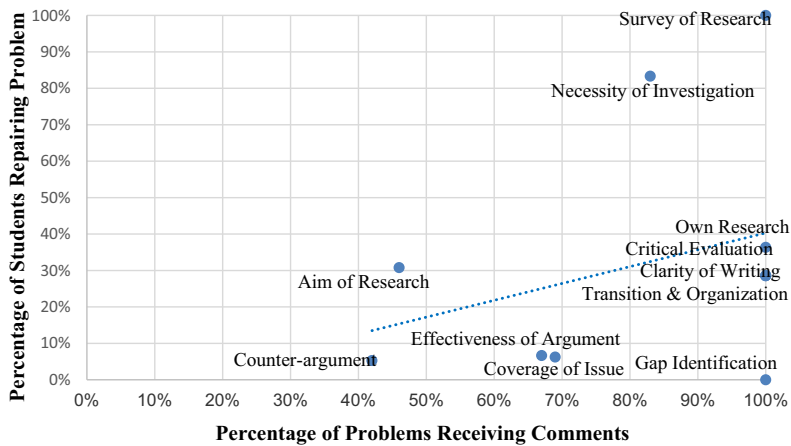


Figure 4. For each substance and high prose issue, relative rates of receiving comments on the issue and relative rates of repairing problems.

was no better than having no comments on an issue in terms of repair rate. Two to four comments improved the likelihood of repair, five or more comments on an issue generally produced the highest likelihood of repair. Receiving more comments appeared to result in higher repair rates, a finding similar to that found in earlier L2 studies regardless of the source of comments (teacher or peer) (Liu and Hansen 2002; Tuzi 2004; Gao et al. 2018). However, 45% of problems received 0–1 comment, 36% received 2–4 comments and only 19% received many comments. Thus, the weaknesses in revision rates can at least be partially explained by the relatively low overlap in issue content among peer comments an author received.

Discussion

Previous research has found that students can help peers in L2 writing (Zhao 2010; Allen and Mills 2016), but it has had little to say about whether and to what extent peer comments can address and help peers solve particular kinds of problems. The present study has shown that, although all problems received some comments, the frequency of comments was not well aligned with the frequency of problems, with the most common problems eliciting fewer comments. Although authors were generally provided with relevant comments from peers, as the prior literature has shown (Yang, Badger, and Yu 2006; Zhao 2010), reviewers were more concerned with ‘basic’ elements (i.e. less problematic issues like *survey of research*) that occurred predominantly in the low group’s writing, while they were less concerned with ‘advanced’ elements (i.e. more problematic issues like *counter-argument*) that appeared in all writing. This indicates that the more ‘advanced’ problems in literature reviews were less frequently detected and were possibly beyond the reviewers’ ability to detect, suggesting that attention from instructors is needed for these elements.

Similar to prior research findings (Nelson and Murphy 1993; Paulus 1999; Tuzi 2004), revision on the basis of peer feedback significantly improved the documents. However, consistent with the reviewers’ focus on ‘basic’ elements, authors predominantly repaired the less challenging problems more. Obviously, authors were somewhat influenced by the focus of peer comments they received (Lundstrom and Baker 2009; Ruegg 2015). However, receiving multiple comments on the same issue was particularly associated with higher repair rates. Unfortunately, only 19% of problems that occurred in the documents received many comments, which at least partially contributed to the relatively low repair rate, especially for the more common problems.

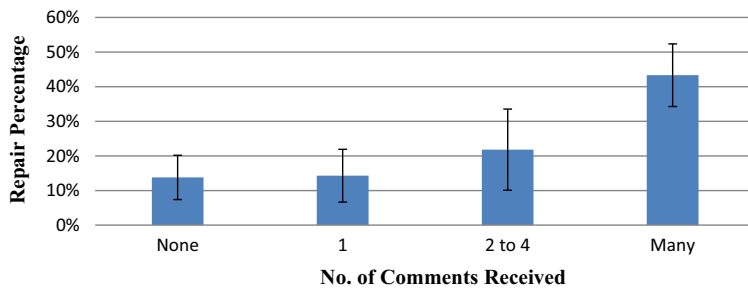


Figure 5. Mean repair rate (and SE bars) as a function of number of comments received on the issue.

The findings in this study suggest that L2 students can help their peers in complex writing tasks and offer scaffolding on substance and higher-level writing issues in their respective zone of proximal development. In particular, high and mid group authors benefited more from peer targeted affordances. Supporting writers of differing abilities is central to instructional concerns in peer assessment (Yu and Lee 2016), and variations in author ability are expected to influence the three main steps/products of peer assessment (initial text, peer feedback and revisions; Nelson and Schunn 2009). As we did not control reviewer ability in this study, this result indicates that, in complicated long essay writings, students with better writing ability (authors of high and mid-rating writings) are more likely to be mediated and assisted by peers, irrespective of reviewer ability. That is to say, as found in Yu and Hu (2016), in mixed-ability groups high-ability students not only contribute to providing feedback, but also benefit from receiving feedback from peers of varying abilities. This may be because they are more capable of making judgments on the received comments or otherwise are more successful as self-regulated learners.

It is noteworthy that peer review appeared to be more useful for certain kinds of feedback as not only more comments were provided, but also more repairs were made on these issues. This implies that, in the complex task of writing literature reviews, EFL learners are more capable of handling some of the more basic elements for which they have already some success and, therefore, can be trusted with these issues. With the more 'advanced' issues in such tasks (that are very difficult for even the strongest students in the class), however, the helpfulness of EFL peer feedback is questionable and should be cautiously monitored. In particular, pre-writing training for both authors and reviewers may need to place emphasis on the more 'advanced' issues so that learners can pay attention to these aspects in both writing and reviewing. Importantly, expert support via lectures, and sometimes one-to-one expert support, should be provided to both writers and reviewers during the peer review process. Note, however, that the current study focuses only on higher-level issues (substance and high prose) in the complex task of writing literature reviews, and, therefore, the results do not suggest that EFL writers can only be trusted to focus on low-level writing issues – relatively few comments were made on low-level writing issues and relatively few revisions were made at this level.

That students repaired most in revision upon receiving multiple comments indicates that, although it is usually perceived that it is the more capable peers who can provide assistance and mediation to peers in learner's zone of proximal development (Vygotsky 1978), the cumulative effect from peers (regardless of level) constitutes more favourable learning conditions. Even though the complexity of the task in writing literature reviews could be too challenging for some students, and therefore they could only repair what was within their limits, receiving the largest amount of comments did afford them to make repairs and improve their drafts. Low group writings repaired most in our study although their repair rate was not statistically correlated with the comments they received. This was probably caused by their having the widest range of problems in their initial drafts and receiving the largest amount of peer feedback, but they could only repair the most 'basic' problems in the writings.

Therefore, to best help peers through feedback, reviewers need to focus more on the problems, especially the more 'advanced' problems in initial texts in complex writing tasks, so that their comments can really count. This is of course a more challenging task for the reviewers, but the benefit they likely will receive from detecting problems and suggesting solutions is expected to be rewarding to both the authors and reviewers (Hu 2005; Min 2005; Rahimi 2013). Therefore, special training and scaffolding for successful L2 peer review (Zhao 2014) in commenting more towards key problems and substance and high prose issues from instructors are needed. To this end, carefully designed feedback procedures, sufficient training, as well as adjustment to the training procedures in the pre-, during- and post-stages of peer feedback (Yu and Lee 2016) are necessary.

In particular, the large distinction between the low-quality group with the other groups in both initial drafts and revision suggests writers of this group need additional support. In the four dimensions of reviewing process in writing (defining a task, detecting a problem, diagnosing a problem and strategy selection; Flower et al. 1986; Hayes et al. 1987), studies have shown difference across author and reviewer abilities. Lower ability authors are less likely to make global revisions because they have a narrow view of the paper and only attempt to correct errors, and they normally avoid diagnosing the problem and use few revision strategies (Patchan and Schunn 2015). The complexity and difficulty in literature review writing, in particular, may be too demanding for low ability authors. Perhaps these students would benefit from instructors using authentic literature review texts as models, allowing students to appreciate the complexity and the kinds of deliberation that are involved in the construction of literature reviews (Kwan 2006). Along with these models, the revised move framework used for coding in this study could provide some useful metalanguage for students and instructors.

Limitations and future work

A few caveats to our findings must be considered. First, the assignment context of this study likely shaped the results. In particular, the focus of feedback was likely strongly influenced by the detailed rubric that focused on substance and high-level writing issues. Without this support, novice writers tend to focus more on lower-level issues (Wallace and Hayes 1991). Also, the online peer review tool required end comments instead of marginalia. This could have affected the focus of the comments because end comments are more likely to focus on substance and high prose issues, whereas marginalia may more likely focus on low prose issues (Patchan et al. 2013).

Another potential limitation involves range restriction. Participants in this study were from the same major of the same university. Their level of English and their domain knowledge were within a particular range. If students were more junior, from less selective universities, or not English majors, different results may have been produced. However, the ability range was large enough to show statistically significant effects in performance even with relatively small numbers.

The current study was text-centric. The study did not consider causes of repair other than received comments. However, repairs could have been made out of various causes, for example, by the author themselves (self-repair) out of self-regulation (Yang and Carless 2013), or from seeing models of good and bad strategies while giving peer feedback. These sources are important areas of study and may require interviews or observations of revision processes.

Studies have shown that reviewer ability and author ability can interact to shape the kinds of comments produced and their effects on revision, at least with L1 writers (Patchan and Schunn 2016). For example, perhaps only comments from high ability reviewers can help high-ability writers with the more difficult kinds of issues. This study did not have sufficient statistical power

to examine such interactions, and future research should further examine this issue in L2 writing contexts.

Disclosure statement

The second author is a co-inventor of the Peerceptiv system.

Funding

This study was supported by The National Social Science Fund of China [18BYY114] .

ORCID

Ying Gao  <http://orcid.org/0000-0003-3565-2351>

Christian Dieter D. Schunn  <http://orcid.org/0000-0003-3589-297X>

References

- Allen, D., and A. Mills. 2016. "The Impact of Second Language Proficiency in Dyadic Peer Feedback." *Language Teaching Research* 20 (4): 498–513.
- Barstow, B., L. Fazio, C. D. Schunn, and K. Ashley. 2017. "Experimental Evidence for Diagramming Benefits in Science Writing." *Instructional Science* 45 (3): 1–20.
- Berggren, J. 2015. "Learning from Giving Feedback: A Study of Secondary-Level Students." *ELT Journal* 69 (1): 58–70.
- Boud, D., and E. Molloy. 2013. "Rethinking Models of Feedback for Learning: The Challenge of Design." *Assessment & Evaluation in Higher Education* 38 (6): 698–712.
- Bruffee, K. A. 1980. *A Short Course in Writing. Practical Rhetoric for Composition Courses, Writing Workshops, and Tutor Training Programs*. Boston, MA: Little, Brown and Company.
- Chang, C. 2016. "Two Decades of Research in L2 Peer Review." *Journal of Writing Research* 8 (1): 81–117.
- Chen, C. W. 2010. "Graduate Students' Self-Reported Perspectives Regarding Peer Feedback and Feedback from Writing Consultants." *Asia Pacific Education Review* 11 (2): 151–158.
- Cho, K., and C. D. Schunn. 2007. "Scaffolded Writing and Rewriting in the Discipline: A Web-Based Reciprocal Peer Review System." *Computers and Education* 48 (3): 409–426.
- Cho, Y. H., and K. Cho. 2011. "Peer Reviewers Learn from Giving Comment." *Instructional Science* 39 (5): 629–643.
- Chong, I. 2017. "How Students' Ability Levels Influence the Relevance and Accuracy of Their Feedback to Peers: A Case Study." *Assessing Writing* 31: 13–23.
- Cote, R. A. 2014. "Peer Feedback in Anonymous Peer Review in an EFL Writing Class in Spain." *Gist Education and Learning Research Journal* 9: 67–87.
- Flower, L., J. R. Hayes, L. Carey, K. Schriver, and J. Stratman. 1986. "Detection, Diagnosis, and the Strategies of Revision." *College Composition and Communication* 37 (1): 16–55.
- Flowerdew, J. 2000. "Discourse Community, Legitimate Peripheral Participation, and the Nonnative-English-Speaking Scholar." *TESOL Quarterly* 34 (1): 127–150.
- Gao, Y., F. Zhang, S. Zhang, and C. D. Schunn. 2018. "Effects of Receiving Peer Feedback in English Writing: A Study Based on Peerceptiv." *Technology Enhanced Foreign Language Education* 180 (2): 3–9, 67.
- Hayes, J. R., L. Flower, K. A. Schriver, J. F. Stratman, and L. Carey. 1987. "Cognitive Processes in Revision." In *Advances in Applied Psycholinguistics. Vol. 2 of Reading, Writing, and Language Learning*, edited by S. Rosenberg, 176–240. New York: Cambridge University Press.
- Hinkel, E. 1997. "Indirectness in L1 and L2 Academic Writing." *Journal of Pragmatics* 27 (3): 361–386.
- Hu, G. 2005. "Using Peer Review with Chinese ESL Student Writers." *Language Teaching Research* 9 (3): 321–342.
- Hu, G., and S. T. E. Lam. 2010. "Issues of Cultural Appropriateness and Pedagogical Efficacy: Exploring Peer Review in a Second Language Writing Class." *Instructional Science* 38 (4): 371–394.
- Huisman, B., N. Saab, J. Driel, and P. Broek. 2018. "Peer Feedback on Academic Writing: Undergraduate Students' Peer Feedback Role, Peer Feedback Perceptions and Essay Performance." *Assessment & Evaluation in Higher Education* 43 (6): 955–968. Advance online publication. doi:10.1080/02602938.2018.1424318.

- Hyland, K., and F. Hyland. 2006. "Feedback on Second Language Students' Writing." *Language Teaching* 39 (2): 83–101.
- Kwan, B. S. C. 2006. "The Schematic Structure of Literature Reviews in Doctoral Theses of Applied Linguistics." *English for Specific Purposes* 25 (1): 30–55.
- Li, H., Y. Xiong, X. Zang, M. L. Kornhaber, Y. Lyu, K. S. Chung, and H. K. Suen. 2016. "Peer Assessment in the Digital Age: A Meta-Analysis Comparing Peer and Teacher ratings." *Assessment & Evaluation in Higher Education* 41 (2): 245–264.
- Liu, J., and J. Hansen. 2002. *Peer Response in Second Language Writing Classroom*. Ann Arbor, MI: University of Michigan Press.
- Liu, N. F., and D. Carless. 2006. "Peer Feedback: The Learning Element of Peer Assessment." *Teaching in Higher Education* 11 (3): 279–290.
- Lundstrom, K., and W. Baker. 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.
- Min, H. T. 2005. "Training Students to Become Successful Peer Reviewers." *System* 33 (2): 293–308.
- Min, H. T. 2008. "Reviewer Stances and Writer Perceptions in EFL Peer Review Training." *English for Specific Purposes* 27 (3): 285–305.
- Nelson, G. L., and J. M. Murphy. 1993. "Peer Response Groups: Do L2 Writers Use Peer Comments in Revising Their Drafts?" *TESOL Quarterly* 27 (1): 135–141.
- Nelson, M. M., and C. D. Schunn. 2009. "The Nature of Feedback: How Different Types of Peer Feedback Affect Writing Performance." *Instructional Science* 37 (4): 375–401.
- Nussbaum, E. M., and G. Schraw. 2007. "Promoting Argument-Counterargument Integration in Students' Writing." *The Journal of Experimental Education* 76 (1): 59–92.
- Patchan, M. M., D. Charney, and C. D. Schunn. 2009. "A Validation Study of Students' End Comments: Comparing Comments by Students, a Writing Instructor, and a Content Instructor." *Journal of Writing Research* 1 (2): 124–152.
- Patchan, M. M., B. Hawk, C. A. Stevens, and C. D. Schunn. 2013. "The Effects of Skill Diversity on Commenting and Revision." *Instructional Science* 41 (2): 381–405.
- Patchan, M. M., and C. D. Schunn. 2015. "Understanding the Benefits of Providing Peer Feedback: How Students Respond to Peers' Texts of Varying Quality." *Instructional Science* 43 (5): 591–614.
- Patchan, M. M., and C. D. Schunn. 2016. "Understanding the Effects of Receiving Peer Feedback for Text Revision: Relations Between Author and Reviewer Ability." *Journal of Writing Research* 8 (2): 227–265.
- Paulus, T. M. 1999. "The Effect of Peer and Teacher Feedback on Student Writing." *Journal of Second Language Writing* 8 (3): 265–289.
- Rahimi, M. 2013. "Is Training Students' Reviewers Worth Its While? A Study of How Training Influences the Quality of Students' Feedback and Writing." *Language Teaching Research* 17 (1): 67–89.
- Ruegg, R. 2015. "The Relative Effects of Peer and Teacher Feedback on Improvement in EFL Students' Writing Ability." *Linguistics and Education* 29: 73–82.
- Schiefele, U. 1999. "Interest and Learning from Text." *Scientific Studies of Reading* 3 (3): 257–279.
- Schunn, C. D. 2016. "Writing to Learn and Learning to Write through SWoRD." In *Adaptive Educational Technologies for Literacy Instruction*, edited by S. A. Crossley and D. S. McNamara. New York, NY: Taylor & Francis.
- Schwarz, B. B., Y. Neuman, J. Gil, and M. Ilya. 2003. "Construction of Collective and Individual Knowledge in Argumentative Activity." *Journal of the Learning Sciences* 12 (2): 219–256.
- Swales, J. M. 2004. *Research Genres: Explorations and Applications*. Cambridge, MA: Cambridge University Press.
- Swales, J. M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge, MA: Cambridge University Press.
- Tsui, A. B. M., and M. Ng. 2000. "Do Secondary L2 Writers Benefit from Peer Comments?" *Journal of Second Language Writing* 9 (2): 147–170.
- Tuzi, F. 2004. "The Impact of e-Feedback on the Revisions of L2 Writers in an Academic Writing Course." *Computers and Composition* 21 (2): 217–235.
- Villamil, O. S., and M. De Guerrero. 2006. "Socio-Cultural Theory: A Framework for Understanding the Socio-Cognitive Dimensions of Peer Feedback." In *Feedback in Second Language Writing: Contexts and Issues*, edited by K. Hyland and F. Hyland, 23–42. New York, NY: Cambridge University Press.
- Vygotsky, L. S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Walker, M. 2015. "The Quality of Written Peer Feedback on Undergraduates' Draft Answers to an Assignment, and the Use Made of the Feedback." *Assessment & Evaluation in Higher Education* 40 (2): 232–247.
- Wallace, D. L., and J. R. Hayes. 1991. "Redefining Revision for Freshmen." *Research in the Teaching of English* 25 (1): 54–66.
- Yang, M., R. Badger, and Z. Yu. 2006. "A Comparative Study of Peer and Teacher Feedback in a Chinese EFL Writing Class." *Journal of Second Language Writing* 15 (3): 179–200.

- Yang, M., and D. Carless. 2013. "The Feedback Triangle and the Enhancement of Dialogic Feedback Processes." *Teaching in Higher Education* 18 (3): 285–297.
- Yu, S., and G. Hu. 2016. "Can Higher-Proficiency L2 Learners Benefit from Working with Lower-Proficiency Partners in Peer Feedback?" *Teaching in Higher Education* 22 (2): 178–192.
- Yu, S., and I. Lee. 2016. "Peer Feedback in Second Language Writing (2005–2014)." *Language Teaching* 49 (4): 461–493.
- 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.
- Zhao, H. 2014. "Investigating Teacher-Supported Peer Assessment for EFL Writing." *ELT Journal* 68 (2): 155–168.
- Zheng, C. 2012. "Understanding the Learning Process of Peer Feedback Activity: An Ethnographic Study of Exploratory Practice." *Language Teaching Research* 16 (1): 109–126.
- Zou, Y, C. D. Schunn, Y. Wang and F. Zhang. 2018. "Student Attitudes That Predict Participation in Peer Assessment." *Assessment & Evaluation in Higher Education* 43 (5): 800–811.