

Can Rating Be More Than a Warm-Up Activity in Classroom Assessments? Exploring the Role of Rating in Peer and Self-Assessments

ECNU Review of Education

1–22

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/20965311251346498

journals.sagepub.com/home/roe**Choo Mui Cheong** (钟竹梅) 

The University of Hong Kong

Xinhua Zhu (祝新华) 

The Hong Kong Polytechnic University

Qi Lu (陆琦) 

Zhejiang University

Abstract

Purpose: This study investigated the relationships between peer rating accuracy, peer feedback, writing revision, and writing improvement as well as those between self-rating accuracy, writing revision, and writing improvement.

Design/Approach/Methods: Data from self- and peer assessment worksheets and the abstract writing task with 114 undergraduate students in Hong Kong were coded into primary variables and then analyzed using structural equation modeling, allowing for a comprehensive exploration of the research purpose.

Findings: The results indicated that peer rating accuracy had a direct positive impact on peer feedback and an indirect impact on students' writing improvement via the mediation of peer feedback and writing revision (i.e., peer rating accuracy → peer feedback → writing revision → writing improvement). Additionally, self-rating accuracy had a direct positive effect on writing revision and

Corresponding author:

Qi Lu, College of Education, Zhejiang University, Hangzhou, China.

Email: 3120103102@zju.edu.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/ham/open-access-at-sage>).

an indirect effect on students' writing improvement through the mediating role of writing revision (i.e., self-rating accuracy → writing revision → writing improvement).

Originality/Value: This study deepens existing understandings of the interplay between rating, feedback, and revision in the assessment process, highlighting their indispensable and mutually reinforcing roles in integrated assessment. Educators can apply this insight to construct more effective assessment processes that promote learning. Implications for enhancing students' rating skills are discussed.

Keywords

Peer feedback, peer rating accuracy, self-rating accuracy, writing improvement, writing revision

Date received: 22 November 2023; revised: 18 January 2024, 18 February 2024; accepted: 14 March 2024

Introduction

In conducting peer and self-assessments, students evaluate the quality of their peers' and their own work, respectively, by providing both quantitative ratings and qualitative feedback (e.g., Cheong et al., 2023; Stančić, 2021; Yan et al., 2022). These two assessment approaches have become widely utilized instructional practices in higher education writing classrooms because they can facilitate Assessment for Learning (AfL) and improve students' academic writing proficiency (e.g., Staberg et al., 2023; Xiang et al., 2022). There is a growing body of research on the qualitative dimensions of peer and self-assessment, including the learning mechanism of providing and receiving peer feedback (e.g., Gao et al., 2023), the importance of various features of peer and self-feedback (e.g., Lu et al., 2021; Wu & Schunn, 2021), and the effectiveness of peer and self-feedback on writing development (e.g., Cheong et al., 2023; Zou et al., 2023). Despite these efforts, there remains a relative paucity of knowledge concerning the value of quality ratings that students could assign to their peers' work as well as their own.

To effectively engage in peer or self-assessment, students must first provide a judgment of their peers' or their own performance, which entails accurately evaluating the quality of work (Chang & de Lemos Coutinho, 2022). Johnson et al. (2009) propose using rating accuracy to assess the quality of this evaluation. Previous research has examined various factors that may impact the accuracy of students' ratings, such as their understanding of rating criteria and lack of prior experience in evaluation (e.g., Carroll, 2020; Rico-Juan et al., 2022). Additionally, longitudinal studies have explored changes in rating accuracy over time and the potential factors that influence these changes (e.g., Birjandi & Siyyari, 2010; Han & Riazi, 2018). Furthermore, previous studies have suggested pedagogical strategies to enhance rating accuracy, including providing specific rating criteria and rubrics and training students on how to use them effectively (e.g., Zhang & Zhang, 2022b).

Despite growing interest in this area, most studies have examined rating accuracy only within the context of the rating activity itself. Very little scholarly attention has been paid to the potential impact of rating accuracy on subsequent student engagement with other integral assessment components such as qualitative feedback provision and writing revision. One concerning issue that arises from this lack of attention is the potential for rating and feedback to be viewed as unconnected activities within the assessment process, with ratings relegated to a perfunctory role in which students simply familiarize themselves with the assessment procedures. However, this perspective may fail to appreciate the interdependent nature of ratings and feedback, which should be viewed as mutually reinforcing components of an integrated assessment process. Furthermore, the possible implications of student rating accuracy for subsequent learning outcomes may be underestimated, despite its potential to impact students' engagement with follow-up feedback and writing revisions.

In light of the aforementioned concerns, it is crucial to explore the role of student rating accuracy in the assessment process, particularly the ways in which peer and self-rating accuracy can affect peer feedback and writing revision and, in turn, ultimately influence students' writing performance. This investigation has the potential to provide educators with a more comprehensive understanding of effectively connecting rating, feedback, and revision activities. Such insights can facilitate more potent peer and self-assessment in classroom contexts, maximizing the educational potential of these assessment approaches.

Literature review

Rating accuracy and feedback in peer assessment and writing improvement

Johnson et al. (2009) suggested using rating accuracy as a measure to evaluate the quality of assessment, which quantitatively compares the scores given by students with those given by their teachers. The teachers' scores are regarded as benchmark ratings for comparison (Han & Riazi, 2018). The smaller the differences between the two ratings, the more precise the students' evaluation. As students assign scores to different objects (i.e., either their peers or themselves), rating accuracy can be further categorized into peer-rating accuracy and self-rating accuracy.

Peer-rating accuracy describes students' ability to accurately evaluate their peers' work (Han & Zhao, 2021; Liu et al., 2019). Several studies have depicted this as rater accuracy (Wang & Engelhard, 2019), underscoring the importance of students' assessment literacy as peer reviewers. According to a criterion-based approach (Han & Riazi, 2018), peer-rating accuracy is operationalized as the difference between peer ratings and teacher ratings (set as criterion ratings). The smaller the difference between these two ratings, the higher the level of peer rating accuracy.

Concerning the association between peer rating accuracy and peer feedback, existing literature offers theoretical support based on the social constructivist perspective. This perspective suggests

that engaging in peer assessment compels students to internalize assessment criteria by requiring them to evaluate their peers' work (Gielen et al., 2011). A deeper understanding, as indicated by the accurate application of criteria, is suggested to potentially lead to higher quality feedback and revision, as both processes require a profound grasp of the assessment criteria (Han & Zhao, 2021; Li et al., 2020).

Notably, while theoretical foundations exist, empirical research has not yet established whether there is a direct relationship between peer rating accuracy and peer feedback. However, Davies (2006, 2009) presents indirect support for their association, revealing students' tendency to provide feedback aligned with their perceived ratings of an essay. Specifically, low-scoring essays attracted negative comments focusing on errors and weaknesses, while high-scoring essays garnered positive feedback with praise. These findings suggest that peer ratings may serve as an epitome of peer feedback. In other words, the accuracy of peer ratings can be considered an indicator of the quality of peer feedback. Therefore, it can be inferred that the accuracy of peer ratings is likely to directly impact peer feedback.

In this study, peer feedback was divided into two distinct components: amount and types. The amount of feedback refers to the number of comments provided by peers (Wu & Schunn, 2021). The types of feedback pertain to the functions that comments serve regarding the text and encompass categories such as summary, praise, problem, and solution (Lu et al., 2021). Notably, "types" in this study denote the total amount of comment types in the feedback rather than the quantity of each feedback type. This integrated perspective aligns with arguments from previous research (e.g., Lu et al., 2021; Mahboob, 2015; Wu & Schunn, 2020), suggesting that focusing solely on individual feedback types, such as solutions, may inhibit the coherent presentation of information necessary for effective student revisions. This study operationalized the construct of peer feedback in terms of amount and types for two primary reasons. First, recent peer feedback research has highlighted these components as having a unique impact on the writing process and product (e.g., Wu & Schunn, 2023). Second, this conceptualization illuminates the tradeoff students face between choosing issues to comment on in greater depth (generating more types) and providing feedback on a greater number of issues (generating a greater amount) (Zong et al., 2021).

In the realm of peer feedback and writing improvement, recent research has confirmed the positive impact of peer feedback on writing quality (e.g., Latifi et al., 2023; Wu & Schunn, 2021). However, when examining the impact of peer feedback on students' writing improvement, the role of writing revision is equally essential. Research consistently indicates that revision serves as an intermediary mechanism linking peer feedback and writing outcomes (Lu et al., 2023; Wu & Schunn, 2021). Peer feedback alerts students to areas requiring improvement, while revision acts as a critical conduit through which students translate feedback into revisions, thereby

enhancing their writing. Therefore, writing revision may play a crucial mediating role between peer feedback and writing improvement.

Based on the aforementioned considerations, we hypothesized that peer rating accuracy influences students' writing improvement through two mediators: peer feedback and writing revisions. This mediating effect may involve a sequential two-stage process, where peer-rating accuracy first influences peer feedback (mediator 1) and writing revision (mediator 2), ultimately affecting students' writing improvement.

Rating accuracy in self-assessment, writing revision, and writing improvement

Self-rating accuracy, akin to peer rating accuracy, reflects a student's ability to accurately evaluate the quality of their own work according to given criteria (Panadero et al., 2016). It is operationalized as the difference between self-ratings and teacher ratings, with a smaller difference indicating higher self-rating accuracy. Examining the nexus between self-rating accuracy and writing revision aligns with the theoretical underpinning of the self-regulated learning perspective. This perspective posits that precise self-assessments fortify metacognitive monitoring and self-reflection skills (Panadero et al., 2016). Accurate self-ratings are usually associated with the ability to identify weaknesses and revision needs (Lu et al., 2024). The potential impact of self-rating accuracy on writing revision and, by extension, writing improvement, becomes apparent.

Prior research has also alluded to a correlation between self-rating ability and learning behaviors in student writers. Skilled self-raters use feedback to reflect on the quality of their writing and consciously apply effective revision strategies, leading to improvements in their writing (e.g., MacArthur, 2018). Although the positive impact of accurate self-evaluations on students' revisions seems intuitive, there is a lack of empirical evidence to support this claim. As Andrade (2019) noted, "a study that closely examines the revisions to writing made by accurate and inaccurate self-assessors, and the resulting outcomes in terms of the quality of their writing, would be most welcome." Thus, our study addresses this gap in the literature by investigating the impact of self-rating accuracy on writing revision and, ultimately, writing improvement.

Overview of the literature and aim of the present study

Through a review of existing literature, this study revealed a limitation in exploring the interconnections between various assessment components that facilitate meaningful learning for students. Two main issues have arisen. First, while previous studies have established the relationship between peer ratings and peer feedback (e.g., Davies, 2009), as well as peer feedback and writing improvement (e.g., Latifi et al., 2023), there is a dearth of research examining the underlying mechanism of how peer rating accuracy impacts students' writing improvement. A deeper

understanding of this mechanism may help educators establish a more cohesive peer assessment process, enabling students to fully benefit from it. Second, the role of self-rating accuracy in self-assessment, particularly its influence on students’ writing revisions, is not well understood in the literature. This knowledge gap may impede the complete understanding of the potential benefits of self-rating activities and the need to improve students’ self-rating accuracy.

Situated in the context of Chinese language learning, this study aims to address the aforementioned issues by investigating the relationships between peer rating accuracy, peer feedback, writing revision, and writing improvement, as well as between self-rating accuracy, writing revision, and writing improvement. By shedding light on the crucial role of rating accuracy in the peer and self-assessment processes, this study’s findings could potentially enhance the effectiveness of these assessment practices for students.

The study posits a hypothesized model, presented in Figure 1. The model proposes that peer rating accuracy influences students’ writing improvement through two sequential mediators: peer feedback and writing revisions. Furthermore, self-rating accuracy was hypothesized to influence students’ writing improvement through its effect on writing revisions. The following research questions guided this study:

- RQ1:** To what extent does peer rating accuracy affect students’ writing improvement through the sequential mediators of peer feedback and writing revision?
- RQ2:** To what extent does self-rating accuracy impact students’ writing improvement through writing revision?

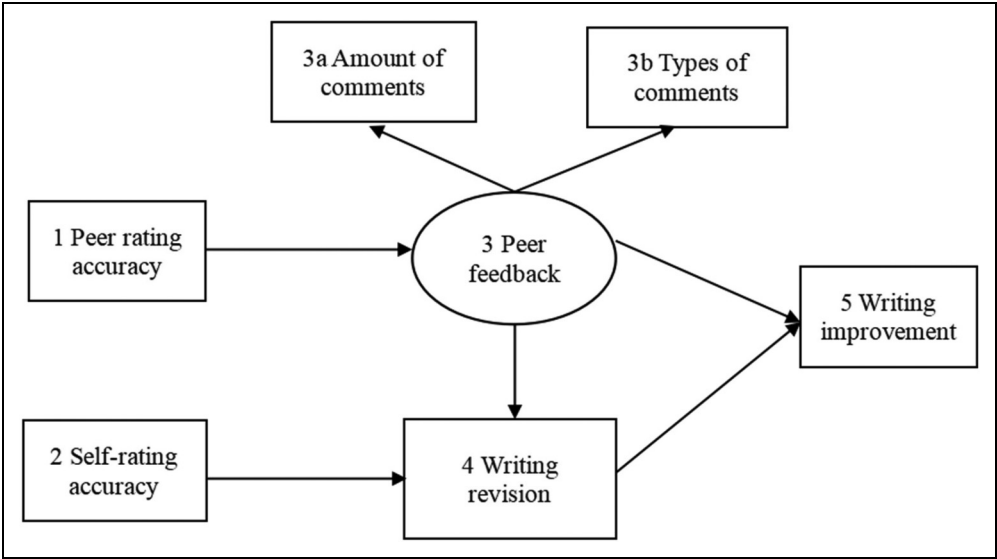


Figure 1. Hypothesized model.

Methods

Contexts and participants

This study is a subset of a broader research endeavor that examines how university-level students acquire knowledge from different forms of in-class assessment, such as self- and peer assessment, in the context of learning Chinese. The participants were 116 BA (Hons.) students (male = 26, female = 90; $M_{\text{age}} = 21.97$, $SD = 1.04$) enrolled in an academic writing course at a Hong Kong university. The course was designed to teach research methods and academic writing skills. The AfL approach has been promoted as a pedagogical device to activate student-centered learning by Hong Kong's Education Bureau since the early 2000s (Curriculum Development Institute, 2004). Therefore, the participants were familiar with peer and self-assessment through frequent classroom use. All students provided informed consent to participate voluntarily. Two students did not complete the peer assessment as they were absent from the peer rating session, and their data were omitted, leaving the data of 114 participants for analysis.

Materials

Abstract writing task and rating criteria. Composing an abstract is an important learning task in academic writing. Students were asked to write an abstract for a selected Chinese academic research article. The article "The challenges in teaching non-Chinese speaking students in Hong Kong Chinese language classrooms" (Kwan, 2014), published in the Newsletter of Chinese Language (later renamed "Current Research in Chinese Linguistics"), was chosen as it was published without an abstract.

The quality of the students' abstract writing was measured on five dimensions: research purpose, method, findings, implications, and language convention (Lu et al., 2021). These criteria are consistent with the "IMRaD" (Introduction, Method, Results, and Discussion) structure generally required for academic writing in the social sciences (Tabuena, 2020) and highlight the importance of readability in a well-structured abstract (Tankó, 2017). Since both competencies to organize the content logically (the first four criteria) and express the content concisely (the last criterion) are important for academic writing, we assigned the same weight to each criterion. Each criterion was rated out of five, all provided with descriptors, with each level ranging across two marks. The highest level was 7–8, which reflected an excellent writing performance, and the lowest level was 0, which indicated an unacceptable writing performance. This design was confirmed and approved by the Curriculum Review Board at the university where this study was conducted.

Peer and self-assessment worksheets. Peer and self-assessment worksheets were designed to help students review their own writing and that of their peers; both instruments provided the same rating

criteria. The self-assessment worksheet required students to score their own performance, while the peer assessment worksheet (see Appendix A) required students to mark their peer’s abstracts, provide justification, and share their opinions for improvement.

Procedure

This procedure was conducted over two lessons in two weeks, as shown in Table 1. During the first lesson, the students attended a lecture on abstract writing and its rating criteria, wrote and submitted an abstract to the course’s online learning platform, and conducted a self-assessment. The total reaction time was 150 min. One week later, students completed a peer assessment in pairs randomly assigned to the platform. They provided both quantitative scores and qualitative comments in 30 min, and later revised their own text according to the received peer assessment in 50 min. Students then submitted their final drafts.

Marking and coding

Abstract writing quality. The first two authors of this article, both with over 10 years of experience in teaching academic writing, marked the students’ first and final drafts based on the rating criteria. Before rating, a meeting was held to ensure that both the markers used the rating criteria accurately and consistently. Sixty first and final drafts were double-marked and the results were reviewed. A mean score was assigned if the gap between the two scores was less than two; otherwise, a score was provided by a third marker (a teacher on the same course), and the score was summed and averaged with the closest score assigned by one of the original markers. The inter-rater reliability for the double-marked data was computed using the intra-class correlation coefficient, 0.88 for the first draft and 0.82 for the final draft, showing good inter-rater reliability. The remaining abstracts were independently marked. The improvement in writing quality was calculated by subtracting the final draft score from the first draft.

Table 1. Procedure of the abstract writing task.

Time	Activity	Procedure
Week 1	1 (50 min)	The course teacher taught students how to write an abstract and explained its rating criteria.
	2 (80 min)	The teacher introduced the writing task, and students drafted the abstract.
	3 (20 min)	The students self-assessed their own work.
Week 2	4 (30 min)	The students assessed their peers’ work and provided scores and comments.
	5 (50 min)	The students revised their abstract based on the received peer assessment.

Rating accuracy. Rating accuracy was calculated in two steps. To compute peer rating accuracy, the score assigned by the markers for the student's first draft was subtracted from the peer rating for the same draft to obtain the difference between the scores. A higher difference score indicated less accurate peer evaluation. However, this association was not intuitive. To simplify the interpretation of the results, a second step was taken to reverse-code the difference score obtained in the first step. Specifically, the maximum score among all the different scores (20.5) was used to subtract the value given to individual drafts calculated in the first step. This reverse-coded score, where a higher value indicated greater peer-rating accuracy, was used in subsequent analyses. For example, if a student initially self-evaluates their first abstract as 30 while the teachers assign a rating of 25, the first step would yield a difference of 5. By subtracting this difference from the maximum value of 20.5, the self-rating accuracy value would amount to 14.5. A higher value represents a more accurate peer evaluation. This same method was used to calculate self-rating accuracy.

Ideas segmentation and coding. The third author and a trained research assistant segmented the peer comments, coded them, and wrote revisions. First, all coders and authors convened to establish consistency in understanding the coding scheme, and trial coding of ten randomly selected sets of peer comments and ten sets of writing revisions was conducted. One coder then continued to code all the remaining sets of peer comments and writing revisions, and the other independently coded 60 sets of randomly selected peer comments and writing revisions. All disagreements were resolved through discussion to minimize coding noise. The inter-rater reliability of the double-coded data was calculated using Cohen's kappa. Idea segmentation and coding were used to analyze the data collected during the assessment process, including the amount and types of comments, as well as the classifications of writing revisions.

Amount of comments. Comments were segmented into separate idea units, because a single comment often conveyed more than one issue. An independent idea unit was defined as stand-alone information that addresses a single aspect of a text (Wu & Schunn, 2020). In total, 1,375 idea units were produced as a result of the segmentation process. The interrater reliability had a kappa value of .78.

Types of comments. Adapted from Wu and Schunn (2020), each idea unit of the peer comment was first coded into the following types (Appendix B): summary (kappa = .82), praise (kappa = .88), problem (kappa = .92), solution (kappa = .84), and suggestion (kappa = .84). Then, the following steps were taken to treat this variable as a continuous variable: (1) counting the quantity of types in every criterion shown in the peer assessment worksheet (Appendix A) and (2) summing up all the types in the five dimensions of criteria. The higher the student's value, the more types of peer comments he or she received.

Writing revision. Microsoft Word’s “Compare Document” tool was used to track the changes between the first and final drafts. Writing revision was coded as the presence or absence of surface revision and meaning revision (Appendix B) (Wu & Schunn, 2021). The difference is that surface revision did not alter the meaning of the original text, but meaning revision did. Format changes were not considered. The surface revision and meaning revision made by each student were added up, and a total of 490 writing revisions were found ($\kappa = .80$).

Results

Descriptive statistics and correlation analysis

The data collected from the marking and coding processes were first entered into SPSS 25.0 for statistical analysis. Descriptive statistics were performed to examine the data distribution characteristics and are shown in Table 2. The mean values for peer rating and self-rating accuracy were 10.25 and 13.54, respectively. Since a higher value indicates a higher rating accuracy, this result indicated that students as self-reviewers make relatively more accurate evaluations. The mean values were 12.06 for the amount of comments and 13.13 for the types of comments. Notably, these two variables have distinct implications, with the former suggesting that students may receive comments that address more issues, while the latter indicates that students may receive comments that elaborate in greater depth. The mean improvement score of students was 3.76. All kurtosis and skewness values were below 3, suggesting normal distributions of the variables (Kline, 2015).

Pearson correlation analysis was then administered to examine the relationship between the primary variables. As expected, peer rating accuracy was positively correlated with peer feedback ($r = .36$ for amount, $r = .45$ for types) and writing revision ($r = .26$); self-rating accuracy was also

Table 2. Descriptive statistics and correlations for the measured variables.

	1	2	3a	3b	4	5
M	10.25	13.54	12.06	13.13	4.30	3.76
SD	4.96	4.96	2.92	2.49	1.78	2.68
Skewness	0.17	−0.33	0.75	−0.01	0.19	1.05
Kurtosis	−0.75	−0.82	0.27	−0.14	−0.88	1.39
1. Peer rating accuracy	1					
2. Self-rating accuracy	0.59**	1				
3a. Amount of comments	0.36**	0.28*	1			
3b. Types of comments	0.45**	0.26**	0.66**	1		
4. Writing revision	0.26**	0.32**	0.26**	0.32**	1	
5. Writing improvement	−0.17	−0.24*	−0.22	0.09	0.29**	1

Note. * $p < .05$; ** $p < .01$.

positively correlated with writing revision ($r = .32$). The correlation between writing revision and writing improvement was 0.29. All of the absolute correlation coefficients between these variables were at a weak to moderate level, ranging from 0.24 to 0.66 at the $p < .05$ or $p < .01$ level.

Structural equation modeling

AMOS 25.0 was employed to further execute structural equation modeling to examine the mechanisms of peer and self-rating accuracy on students' writing improvement. According to Hu and Bentler (1999), the following indices were used to examine the model data fit: comparative fit index (CFI; good > 0.90), Tucker Lewis index (TLI; good > 0.90), and root mean square error of approximation (RMSEA; acceptable < 0.08). Figure 2 shows the standardized coefficient estimates for the final path model. The model fit the data well [$\chi^2(6) = 11.99$, $\chi^2/df = 2.00$, CFI = 0.97, TLI = 0.92, RMSEA = 0.08]. The model accounted for 25.2% of the variance in peer feedback, 16.4% of the variance in writing revision, and 9% of the variance in students' writing improvement.

Concerning RQ1, a bootstrapping procedure with a bias-corrected confidence interval (95%) was used to test the significance of the mediating effects in the model. The mediating effects were assumed to be significant when zero was beyond the confidence interval (MacKinnon et al., 2004). The results showed that peer rating accuracy exerted a positive effect on peer feedback ($\beta = .50$, $p = .002$), whereas no significant effect of peer feedback on students' writing improvement was found. Peer comments alone could not mediate the relationship between peer rating accuracy and writing improvement ($\beta = .01$, 95% CI $[-0.10, 0.09]$, $p = .75$). Moreover, peer rating accuracy

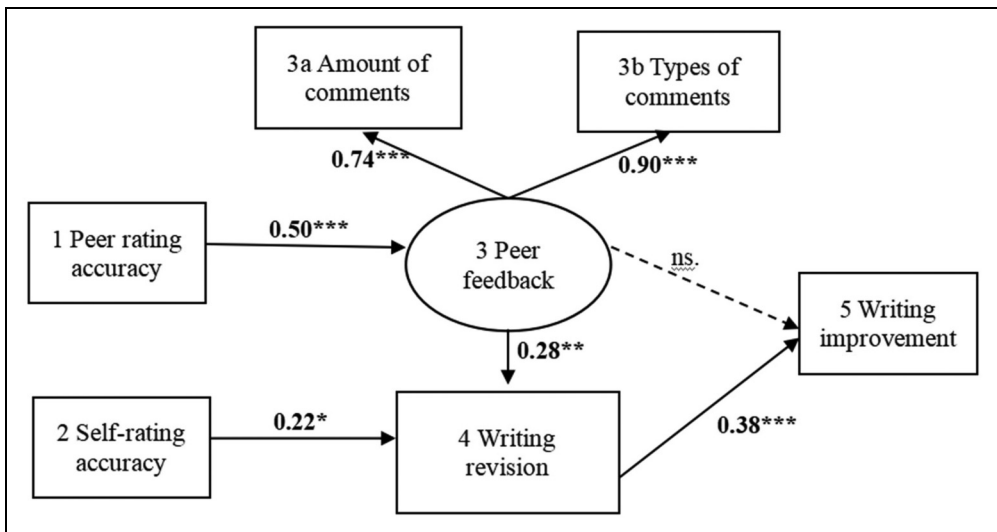


Figure 2. Final path model with standardized coefficient estimates.

Note. *** $p < .001$, ** $p < .01$, * $p < .05$, ns = not significant (dotted line).

can indirectly affect ($\beta = .05$, 95% CI [0.02, 0.14], $p = .003$) students' writing improvement via the mediators of peer comments and writing revisions. With regard to RQ2, self-rating accuracy was found to have a positive effect on writing revisions ($\beta = .22$, $p = .038$). Particularly, writing revisions mediated the relationship between self-rating accuracy and writing improvement ($\beta = .08$, 95% CI [0.02, 0.17], $p = .013$).

Discussion

This study aimed to investigate how peer rating accuracy works together with peer feedback and writing revision to affect students' writing improvement and how self-rating accuracy works with writing revision to affect the improvement outcome in a Chinese language learning context. The major findings are discussed below.

Effects of rating accuracy in peer and self-assessment processes

The structural equation modeling (SEM) results demonstrated that rating accuracy exerted a significant positive influence on both peer feedback and writing revision. More specifically, peer rating accuracy had a direct positive effect on peer feedback, which in turn had a positive impact on writing revision and ultimately led to writing improvement. Similarly, self-rating accuracy had a direct positive effect on writing revision, which contributed to writing improvement. The results could be interpreted in two ways. First, students were able to align their quantitative evaluations with the qualitative feedback provided to their peers. This finding aligns with those of Davies (2009), who asserted that students were consistent in their performance when providing ratings and feedback. Second, students may benefit from a combination of ratings and feedback. This is in line with the findings of Li et al. (2020), who reported that the effect size on learning was smaller when feedback alone was provided ($g = .176$, $p < .05$) compared to when both ratings and feedback were provided ($g = .349$, $p < .001$). These findings suggest that ratings can be used to supplement feedback.

Furthermore, they suggest that the utilization of ratings can serve a dual purpose by not only acting as a warm-up exercise to enhance students' familiarity and sense of ownership in classroom assessment but also by yielding significant impacts on their learning. Given that each component (i.e., rating, feedback, and revision) of the assessment is intimately linked to students' comprehension, internalization, and utilization of the assessment criteria (Han & Zhao, 2021), rating accuracy may play a crucial role in determining whether students can fully engage in the assessment and derive its full benefits (e.g., learning from rating, feedback, and revision). Additionally, as Han and Zhao (2021) have argued, successfully providing accurate ratings can promote students' perception that such practices are meaningful to their learning and stimulate their participation in the assessment.

The mediating effect of writing revision between peer and self-rating accuracies and writing improvement

The SEM results demonstrated that peer feedback can function in tandem with writing revision to enact a dual-stage mediation process for writing improvement. Specifically, peer rating accuracy was found to positively impact feedback, which in turn motivated students to integrate the feedback into their revisions, ultimately leading to writing improvement. Considering the insignificant mediating effect of peer feedback as a single mediator, writing revision appears to play a pivotal role in amplifying the effect of peer feedback on writing improvement. The revision process not only links peer assessment with its outcome, but also embodies the process of interaction, negotiation, and co-construction of meaning among peers (Wang & Lee, 2021). That is to say, students are more likely to incorporate feedback provided by peers into their revisions only when they successfully engage in this “dialogue” and recognize the feedback as valuable. Notably, participating in this “dialogue” can further foster students’ metacognitive development by promoting their reflection on how to effectively use language to construct meaning (known as “linguaging”) (Zhang & Zhang, 2022a).

Furthermore, students with high rating accuracy tended to be more sensitive to numerical scores and exhibited a strong performance-approach goal orientation, indicating a desire to achieve high scores. Consequently, these students were more likely to take advantage of the information provided by their peers, including both ratings and feedback (Cheong et al., 2023), and actively incorporate it into their writing revisions to improve their scores. This may also explain that both peer and self-rating accuracies inevitably affect students’ writing improvement via writing revision.

Pedagogical suggestions

The research findings offer some pedagogical insights for the writing classroom in higher education. First, the ratings are recommended to be conducted as regular activities in peer and self-assessment rather than as only warm-up exercises, as combining rating with feedback promotes more student learning (e.g., Li et al., 2020; Middleton et al., 2023). Second, students should be trained or given explicit instruction on how to provide an accurate judgment of the quality of their own work and that of their peers. Since the evaluation process is closely associated with the use of assessment criteria, teachers could conduct an exemplar analysis in the use of the essay’s criteria and create opportunities for independent practice (Li et al., 2016). Moreover, providing timely guidance and feedback on students’ accuracy in peer and self-rating could raise their awareness as peer and self-reviewers, especially when students are novices in evaluation (Siyyari & Ghorban Daei, 2016). Time spent at this initial stage is worthwhile, as the rippling effect will benefit the students in subsequent engagement in the assessment. Lastly, as the current study showed that writing revision played a vital mediating role, teachers could

encourage students to follow up on feedback by constructively and critically facilitating inner reflection during revisions (Zhang & Hyland, 2022) as well as by requiring students to cross-check the feedback against their revisions and, subsequently, their writing outcomes.

Conclusions

This study contributes to the literature in three ways. First, it extends the existing research on the role of rating accuracy in the assessment process by distinguishing between the different roles of peer and self-rating accuracy. Specifically, it provides empirical evidence that peer rating accuracy fosters peer feedback and self-rating accuracy promotes writing revision. This finding deepens our understanding of the connected nature of rating and feedback, which together constitute a more cohesive peer assessment process. Furthermore, it advances knowledge regarding the impact of rating on students' engagement with follow-up feedback and their revision practices. Second, while previous research has extensively explored the relationship between feedback and revision, the link between rating and revision has received less attention. This study addresses this gap and highlights writing revision as a key factor connecting the roles of peer and self-rating accuracy in improving writing outcomes. The findings provide a more comprehensive understanding of how students can optimize the benefits of rating, feedback, and revisions—critical components of the assessment process—to enhance their writing learning.

There are a few limitations worth mentioning. First, we did not elucidate the source of revisions. Although students mainly revised their work based on the peer feedback they received, they also completed extra revisions motivated by new thoughts generated from reflecting on peer feedback (e.g., Pham, 2022). Indeed, peer rating accuracy may be an important factor influencing the source of revisions as students may form an initial judgment on the usefulness of peer feedback according to the quality of the peer rating. Future research could explore the relationship between peer rating accuracy, peer feedback, and the source of writing revisions, which would provide valuable information on how to promote students' use of peer assessment. Second, the results were primarily quantitative and could be complemented with more qualitative data, such as semi-structured interviews, to further delve into students' processes for conducting and using peer and self-ratings.

Authors' note

Qi Lu has recently changed her affiliation to The Chinese University of Hong Kong.

Contributorship

Qi Lu, Xinhua Zhu, and Choo Mui Cheong designed the methodologies of the research project. Qi Lu and Xinhua Zhu collected and analyzed the data collaboratively. Qi Lu drafted the initial manuscript (including

the figures and tables). Choo Mui Cheong and Xinhua Zhu later elaborated and revised the manuscript while maintaining close communication with Qi Lu. All authors contributed to the article and approved the submitted version.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Ethical statement

The studies involving human participants were reviewed and approved by the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University. All the participants gave consent to participate voluntarily.


Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially supported by a grant from the Joint Supervision Scheme with the Chinese Mainland, Taiwan and Macao Universities - Zhejiang University [Project ID P0031773].

ORCID iDs

Choo Mui Cheong  <https://orcid.org/0000-0001-9590-7681>

Xinhua Zhu  <https://orcid.org/0000-0003-2179-8691>

Qi Lu  <https://orcid.org/0000-0001-7926-4364>

References

- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, <https://doi.org/10.3389/feduc.2019.00087>
- Birjandi, P., & Siyyari, M. (2010). Self-assessment and peer-assessment: A comparative study of their effect on writing performance and rating accuracy. *Iranian Journal of Applied Linguistics (IJAL)*, 13(1), 23–45.
- Carroll, D. (2020). Observations of student accuracy in criteria-based self-assessment. *Assessment & Evaluation in Higher Education*, 45(8), 1088–1105. <https://doi.org/10.1080/02602938.2020.1727411>
- Chang, K., & de Lemos Coutinho, L. (2022). Administering homework self-grading for an engineering course. *College Teaching*, 70(3), 387–395. <https://doi.org/10.1080/87567555.2021.1957666>
- Cheong, C. M., Luo, N., Zhu, X., Lu, Q., & Wei, W. (2023). Self-assessment complements peer assessment for undergraduate students in an academic writing task. *Assessment & Evaluation in Higher Education*, 48(1), 135–148. <https://doi.org/10.1080/02602938.2022.2069225>
- Curriculum Development Institute. (2004). *Promoting assessment for learning in English language education*. Hong Kong Government Printer.

- Davies, P. (2006). Peer assessment: Judging the quality of students' work by comments rather than marks. *Innovations in Education and Teaching International*, 43(1), 69–82. <https://doi.org/10.1080/14703290500467566>
- Davies, P. (2009). Review and reward within the computerised peer-assessment of essays. *Assessment & Evaluation in Higher Education*, 34(3), 321–333. <https://doi.org/10.1080/02602930802071072>
- Gao, Y., An, Q., & Schunn, C. D. (2023). The bilateral benefits of providing and receiving peer feedback in academic writing across varying L2 proficiency. *Studies in Educational Evaluation*, 77, 101252. <https://doi.org/10.1016/j.stueduc.2023.101252>
- Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education*, 36(6), 719–735. <https://doi.org/10.1080/03075071003759037>
- Han, C., & Riazi, M. (2018). The accuracy of student self-assessments of English-Chinese bidirectional interpretation: A longitudinal quantitative study. *Assessment & Evaluation in Higher Education*, 43(3), 386–398. <https://doi.org/10.1080/02602938.2017.1353062>
- Han, C., & Zhao, X. (2021). Accuracy of peer ratings on the quality of spoken-language interpreting. *Assessment & Evaluation in Higher Education*, 46(8), 1299–1313. <https://doi.org/10.1080/02602938.2020.1855624>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Johnson, R. L., Penny, J. A., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance tasks*. The Guilford Press.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Kwan, C. Y. (2014). The challenges in teaching non-Chinese speaking students in Hong Kong Chinese language classrooms [in Chinese]. *Newsletter of Chinese Language*, 93(1), 39–57.
- Latifi, S., Noroozi, O., & Talaei, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, 31(2), 655–669. <https://doi.org/10.1080/10494820.2020.1799032>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193–211. <https://doi.org/10.1080/02602938.2019.1620679>
- Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245–264. <https://doi.org/10.1080/02602938.2014.999746>
- Liu, J., Guo, X., Gao, R., Fram, P., Ling, Y., Zhang, H., & Wang, J. (2019). Students' learning outcomes and peer rating accuracy in compulsory and voluntary online peer assessment. *Assessment & Evaluation in Higher Education*, 44(6), 835–847. <https://doi.org/10.1080/02602938.2018.1542659>
- Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance. *Assessing Writing*, 56, 100725. <https://doi.org/10.1016/j.asw.2023.100725>
- Lu, Q., Yao, Y., & Zhu, X. (2024). The effects of writing revisions on writing improvement among undergraduate students: The moderating role of self-rating accuracy. *The Asia-Pacific Education Researcher*, 33(3), 535–544. <https://doi.org/10.1007/s40299-023-00749-7>

- Lu, Q., Zhu, X., & Cheong, C. M. (2021). Understanding the difference between self-feedback and peer feedback: A comparative study of their effects on undergraduate students' writing improvement. *Frontiers in Psychology, 12*, 739962. <https://doi.org/10.3389/fpsyg.2021.739962>
- MacArthur, C. A. (2018). Evaluation and revision. In S. Graham, C. A. MacArthur, & M. A. Hebert (Eds.), *Best practices in writing instruction* (pp. 287–308). Guilford Publications.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4
- Mahboob, A. (2015). Understanding and providing 'cohesive' and 'coherent' feedback on writing. *Writing & Pedagogy, 7*(2), 401–422. <https://doi.org/10.1558/wap.v7i2-3.26461>
- Middleton, T., Ahmed Shafi, A., Millican, R., & Templeton, S. (2023). Developing effective assessment feedback: Academic buoyancy and the relational dimensions of feedback. *Teaching in Higher Education, 28*(1), 118–135. <https://doi.org/10.1080/13562517.2020.1777397>
- Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*, 803–830. <https://doi.org/10.1007/s10648-015-9350-2>
- Pham, H. T. P. (2022). Computer-mediated and face-to-face peer feedback: Student feedback and revision in EFL writing. *Computer Assisted Language Learning, 35*(9), 2112–2147. <https://doi.org/10.1080/09588221.2020.1868530>
- Rico-Juan, J. R., Cachero, C., & Macià, H. (2022). Influence of individual versus collaborative peer assessment on score accuracy and learning outcomes in higher education: An empirical study. *Assessment & Evaluation in Higher Education, 47*(4), 570–587. <https://doi.org/10.1080/02602938.2021.1955090>
- Siyyari, M., & Ghorban Daei, S. (2016). The effect of formative and summative self-assessment of writing on the writing performance and self-rating accuracy of Iranian EFL learners. *Foreign Language Research Journal, 6*(2), 343–364.
- Staberg, R. L., Febri, M. I. M., Gjøvik, Ø, Sikko, S. A., & Pepin, B. (2023). Science teachers' interactions with resources for formative assessment purposes. *Educational Assessment, Evaluation and Accountability, 35*(1), 5–35. <https://doi.org/10.1007/s11092-022-09401-2>
- Stančić, M. (2021). Peer assessment as a learning and self-assessment tool: A look inside the black box. *Assessment & Evaluation in Higher Education, 46*(6), 852–864. <https://doi.org/10.1080/02602938.2020.1828267>
- Tabuena, A. C. (2020). Students' perception in the implementation of the IMRaD structure approach and its implications on the research writing process. *International Journal of Research Studies in Education, 9*(7), 55–65. <https://doi.org/10.5861/ijrse.2020.5913>
- Tankó, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes, 27*, 42–55. <https://doi.org/10.1016/j.jeap.2017.04.003>
- Wang, J., & Engelhard, G., Jr. (2019). Conceptualizing rater judgments and rating processes for rater-mediated assessments. *Journal of Educational Measurement, 56*(3), 582–609. <https://doi.org/10.1111/jedm.12226>
- Wang, L., & Lee, I. (2021). L2 learners' agentic engagement in an assessment as learning-focused writing classroom. *Assessing Writing, 50*, 100571. <https://doi.org/10.1016/j.asw.2021.100571>

- Wu, Y., & Schunn, C. D. (2020). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology*, 60, 101826. <https://doi.org/10.1016/j.cedpsych.2019.101826>
- Wu, Y., & Schunn, C. D. (2021). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal*, 58(3), 492–526. <https://doi.org/10.3102/0002831220945266>
- Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology*, 73, 102160. <https://doi.org/10.1016/j.cedpsych.2023.102160>
- Xiang, X., Yuan, R., & Yu, B. (2022). Implementing assessment as learning in the L2 writing classroom: A Chinese case. *Assessment & Evaluation in Higher Education*, 47(5), 727–741. <https://doi.org/10.1080/02602938.2021.1965539>
- Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A pairwise and network meta-analysis. *Educational Research Review*, 100484. <https://doi.org/10.1016/j.edurev.2022.100484>
- Zhang, Z. V., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, 100586. <https://doi.org/10.1016/j.asw.2021.100586>
- Zhang, J., & Zhang, L. J. (2022a). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning*, 1–26. <https://doi.org/10.1080/09588221.2022.2069822>
- Zhang, X. S., & Zhang, L. J. (2022b). Sustaining learners' writing development: Effects of using self-assessment on their foreign language writing performance and rating accuracy. *Sustainability*, 14(22), 14686. <https://doi.org/10.3390/su142214686>
- Zong, Z., Schunn, C. D., & Wang, Y. (2021). What aspects of online peer feedback robustly predict growth in students' task performance? *Computers in Human Behavior*, 106924. <https://doi.org/10.1016/j.chb.2021.106924>
- Zou, D., Xie, H., & Wang, F. L. (2023). Effects of technology enhanced peer, teacher and self-feedback on students' collaborative writing, critical thinking tendency and engagement in learning. *Journal of Computing in Higher Education*, 35(1), 166–185. <https://doi.org/10.1007/s12528-022-09337-y>

Appendix A. Peer assessment worksheet.

Assessment criteria		Peer assessment				
		8	6	4	2	0
Research purpose		The purpose of the study is stated clearly, concisely, and accurately, suggesting that the student understands the original article perfectly.	The purpose is stated clearly, suggesting that the student reasonably understands the original article.	The purpose is stated but not clearly, suggesting that the student only partially understands the original article.	The purpose stated is inappropriate, suggesting that the student hardly understands the original article.	The purpose is either missing or totally irrelevant.
Research method		Key issues in the method (e.g., sample, procedure and data analysis) are presented thoroughly, clearly, and concisely.	Most key issues in the method are clearly presented.	One key issue in the method remains missing while the others are clearly described.	Two or more issues in the method are missing.	The method is missing or totally irrelevant.

(continued)

Appendix A. (continued)

Assessment criteria		Peer assessment				
		Justification for the mark (i.e., What do you think of your peer's composition?)				
		8	6	4	2	0
Research findings	The abstract conveys all key findings clearly and concisely.	All key findings are included but the presentation is not concise.	One secondary finding is missing; the presentation needs improvement in clarity and conciseness.	One or more primary findings are missing.	The main findings are missing or totally irrelevant.	
Implications	An implication derives logically from the main findings.	An implication is included and somewhat related to the findings.	An implication is present. However, it is either not clear enough or not tied to the findings.	Although an implication is present, it is incomplete, partially incorrect, or not tied to the findings.	No implication is present or the implication is totally irrelevant.	
Language convention	The language used is coherent, objective, concise,	The language used is objective, fluent, and appropriate but	The language used is objective and mostly grammatical.	While the language used roughly conveyed the author's meaning,	The expressions and sentences	

(continued)

Appendix A. (continued)

Assessment criteria		Peer assessment	
		Justification for the mark (i.e., What do you think of your peer's composition?)	Suggestion(s) for improvement
8	and accurate, with sufficient lexical variation.	6	0
		with occasional errors.	hardly convey logical meaning.
		However, there are grammatically flawed sentences and inappropriate expressions.	The sentences are replete with inappropriate expressions, which makes the abstract difficult to read.
		4	2
		2	0
		0	8

Appendix B. Coding scheme for peer comments and writing revisions.

Category	Definition	Example
<i>Peer comments</i>		
Summary	Comments on what the author wrote	You summarized the difficulties encountered by non-Chinese speaking students in learning Chinese in Hong Kong.
Praise	Comments that indicate good features of the current writing	Your abstract was well written.
Problem	Comments that point out errors in the current writing	Some of the research findings you mentioned were not clear enough.
Solution	Explicit advice with a specific description of how to resolve the problems or improve the quality of the current writing	It would be better to state that the uneven quality of teachers was caused by inappropriate assessments for teachers.
Suggestion	General or heuristic advice on how to resolve the problems or improve the quality of the current writing	You should use more concise expressions.
<i>Writing revision</i>		
Surface revision	Changes that did not alter the meaning of the original text	<p><i>First draft:</i> Hong Kong's educational institutions should focus on these issues.</p> <p><i>Peer comments:</i> It is not clear what you mean by "issues." I suggest you add "discussed above."</p> <p><i>Final draft:</i> Hong Kong's educational institutions should focus on the issues discussed above.</p>
Meaning revision	Changes that altered the meaning of the original text	<p><i>First draft:</i> In sum, there are still several challenges in teaching non-Chinese speaking students in Hong Kong.</p> <p><i>Peer comments:</i> I suggest you also state what should be done to deal with this situation.</p> <p><i>Final draft:</i> In sum, there are still several challenges in teaching non-Chinese speaking students in Hong Kong. Relevant educational departments should pay more attention to these challenges and find appropriate solutions.</p>