

Rubric co-creation to promote quality, interactivity, and uptake of peer feedback

Da Yan^{a*}

*^aSchool of Foreign Languages, Xinyang Agriculture and Forestry University,
Xinyang, China*

Address: No. 1 Beihuan Road, Yang Shan District, Xinyang, Henan, 465000, China.

Email*: alexyan1987@outlook.com

Orcid: <https://orcid.org/0000-0002-1265-9772>

Rubric co-creation to promote quality, interactivity, and uptake of peer feedback

The aim of the study was to examine the effects of rubric co-creation on peer feedback. From a social constructivist perspective, rubric co-creation might have the ability to promote the quality of feedback messages, interactivities within feedback processes, and uptake of feedback information in a peer-based collaborative setting of higher education. Two groups of English Language Studies majors (N=57) from a Chinese university were recruited. In a writing class enacting peer feedback, the experimental group participated in the co-creation of rubrics whereas the control group accepted and used the co-created rubrics. Results showed that rubric co-creation had significantly positive effects on students' use of criteria and the production of meaningful feedback messages instead of feedback writing styles. Regarding interaction patterns, the experimental group established active and bi-directional dialogic communication between peer learners. For uptake of feedback, the experimental group made significantly more critical application of feedback messages compared to the control group. The paper has implications for the enactment of rubric co-creation in peer-based learning and the promotion of learner agency in dialogic feedback interactions.

Keywords: peer feedback; rubric; co-creation; assessment criteria

Introduction

Feedback has long been regarded as a crucial component of teaching and learning (Hattie and Timperley 2007). In recent years, the research focus on feedback has shifted from a traditional view, in which the students were passive recipient of feedback information provided by teachers, peers, and/or technology-based tools, to a

learner-centered paradigm, where feedback is an interactive and constructive process prioritizing student agency, engagement, and uptake (Carless 2015; Winstone and Carless 2019). A prominent example is the wide application of peer feedback in contemporary higher education to enhance learning processes and outcomes (Kerman, Banihashem, and Noroozi 2023). Literature has endorsed its affordances in: (1) alleviating teacher burden for delivering feedback at scale (Nicol 2010); (2) constructing a dialogic environment supporting collaborative learning and evaluative capabilities (Boud and Molloy 2013; Er, Dimitriadis, and Gašević 2021); and (3) engaging students with learner-centered feedback (To 2022).

Noteworthy, peer feedback is a complex and demanding process (Hamer et al. 2015), where students lacking prerequisite knowledge and feedback literacy need social support and cognitive scaffolding (Xu and Carless 2017). Scholars have acknowledged that including students in rubric co-construction resulted in a better understanding of the assessment criteria and performance in using feedback (Panadero and Romero 2014). However, rubric co-creation remains scantily practiced and understudied in peer feedback (Fraile, Panadero, and Pardo 2017), leaving a dearth of knowledge regarding its effects on peer-based collaborative learning (Taylor, Kisby, and Reedy 2024). The issue merits closer scrutiny of the efficacies, interactivity, and outcomes of peer feedback, against the growing awareness that peer feedback is a formative assessment practice (Panadero, Jonsson, and Alqassab 2018) featuring

dialogic interaction (Zhao et al. 2023) for collaborative sense-making and meaning-negotiation (Zhu and Carless 2018).

Addressing Challenges Facing Peer Feedback

The core value of peer feedback is the provision of qualitative information about peer performance, with which learning improvement might be formed (Panadero, Jonsson, and Alqassab 2018). However, the implementation of peer feedback does not necessarily transform into enhancements in learning improvement and engagement, unless the challenges it faces are addressed properly. Extant literature indicates that the success of peer feedback is hindered by unsatisfactory quality (Kerman, Banihashem, and Noroozi 2023). The issue is influenced by individual and contextual factors. First, the interactivity and diversity within peer feedback activities call for a high level of feedback literacy and prior knowledge (Han and Xu 2020).

Unfortunately, many studies have argued that students failed to properly produce meaningful feedback because of limited higher-order thinking abilities (Er, Dimitriadis, and Gašević 2021; Zhu and Carless 2018). Second, acceptance of peer feedback information is also influenced by a set of affective factors, e.g., socio-affective readiness, trust, disposition, and attitude (Huisman et al. 2018). Particularly, distrust and low tolerance of criticality have been regarded as major obstacles impeding effective peer feedback enactments (Kerman et al. 2024).

Correspondingly, researchers and educators have made efforts to support the quality of peer feedback. Two research trends could be synthesized from the literature: 1) providing training or instructional scaffoldings to cultivate feedback literate learners; and 2) offering technological or material support in peer feedback practices. For example, Han and Xu (2020) used teacher feedback as follow-up responses and scaffoldings to peer feedback in a master level academic writing course. The research showcased that support from teachers would cognitively and socio-affectively enhance feedback literacy and peer feedback performance. Similarly, Man, Kong, and Chau (2022) documented the effects of peer review training on feedback literacy development of four focal participants recruited from a Chinese university. After the five-session interventional activities including briefing, modeling, and evaluating training, the participants reported growth in epistemological, affective, and practical aspects of feedback literacy. Apart from these instructional scaffoldings, efforts have been made to afford students with technological and material support in peer feedback. In a recent study by Wood (2022), technology-mediated platforms were applied to promote dialogic peer feedback of 14 undergraduate students from a Korean university. The research showed that students attained cognitive, socio-affective, and relational engagement in processing and adopting peer feedback in addition to the construction of an interactive and collaborative learning community. Similarly, Gong and Yan (2023) have utilized danmaku, or the instant on-screen comments, to elicit synchronous peer feedback in

an L2 oral classroom. Regarding material support in peer feedback, Ashton and Davies (2015) compared the effects of scaffolded rubrics against traditional ones in a MOOC writing course. The result indicated that scaffolded rubrics led to better writing performance, particularly in subjective, complex, and unfamiliar items. Similarly, Gielen and De Wever (2015) applied a feedback template inclusive of multiple structural complexities in a wiki-based setting. Research findings revealed that the procedural facilitator exerted significant effects on feedback quality and product scores.

In sum, the major challenges facing peer feedback practices stem from the cognitive and socio-affective domains of collaborative learning (Nelson and Schunn 2009). Although support has been provided to improve feedback literacy and quality, there is still a paucity of efforts enabling students to become ‘owners of their own learning’ (Black and Wiliam 2009, 8) and promoting agentic peer feedback processes (Er, Dimitriadis, and Gašević 2021; Nieminen et al. 2022). A possible solution may be the enactment of rubric co-creation, an elevation of learners from criteria users to co-developers (Fraile, Panadero, and Pardo 2017).

Rubric Co-creation as a Support for Feedback

According to Brookhart (2018, 1), a rubric ‘articulates expectations for student work by listing criteria for the work and performance level descriptions across a continuum of quality’. As a comprehensively applied assessment instrument, rubric delivers: (1)

standardization of evaluation processes (Reddy and Andrade 2010); (2) students' better understanding of the assessment criteria and cognizance of teachers' expectations of assessment (Kilgour et al. 2020); and (3) augmented performance, motivation, and self-regulation (Brookhart and Chen 2015).

Historically, rubrics have been applied mainly as rating instruments (Popham 1997). However, its use in summative assessment situations such as high stakes testing provokes criticism, e.g., validity issues, bias in assessment, instrumentalism or 'criteria compliance' among learners, and trivial impact on learning caused by superficial implementation (Panadero and Jonsson 2020). Consequently, scholars advocate formative and learner-centered use of rubric to support student learning and instructional improvement, e.g., making assessment criteria explicit, and sharing rubric with students (Panadero and Jonsson 2013). Furthermore, it is argued that students' involvement in rubric creation could promote their autonomy and understanding of assessment criteria (Panadero and Romero 2014), which would in turn foster their abilities to use and ownership of the assessment criteria (Kilgour et al. 2020).

The seminal work by Fraile, Panadero, and Pardo (2017) explored the effects of rubric co-creation on performance, self-regulation, and self-efficacy of Sport Science majors enrolled in a Spanish university. All participants in the quasi-experimental study used the co-created rubric while only the experimental group members took part in the co-creation processes. Data collected from two self-reported

questionnaires showed that there were no significant between-groups differences regarding self-efficacy and self-regulation whereas the results emerged from the thinking-aloud protocol unveiled a higher level of self-regulation in the experimental group. Additionally, the experimental group outperformed the control group in only one of the three tasks measured in the study. The work has established a new venue for follow-up research on rubric co-creation. Santana-Martel and Perez-Garcías (2022) continued to examine the relationship between rubric co-creation, agency, and self-regulation in a Spanish public university. The multiple case study indicated that students reaped comprehensive gains in regulatory and learning skills from the co-creation processes. Zhao, Zhou, and Dawson (2021) discovered the potential of rubric co-creation in promoting attitudes and motivations in a business program. Furthermore, a few other studies attempted to ameliorate the strategies and processes for rubric co-creation reported by Fraile, Panadero, and Pardo (2017). For example, Bacchus et al. (2020) sought to supplement the co-created rubric with exemplars, finding that the combination has yielded better outcomes regarding assessment transparency and affective engagement. In a similar fashion, Kilgour et al. (2020) reported an eight-step procedure for co-creating rubric, whose outcomes not only broaden students' understanding of assessment criteria but also promote transparency and objectivity in assessment processes. Furthermore, Kilgour et al. (2020) acknowledged the potential weaknesses of rubric co-creation: (1) time constraints for the construction processes; and (2) risk of causing view clashes and disagreements.

In summary, the reviewed studies consolidated our understanding of rubric co-creation, particularly its effects on performance and self-regulation. However, looking retrospectively at Fraile, Panadero, and Pardo (2017), their calls for future research remained largely unanswered. First, there were insufficient interventional studies exploring the effects of rubric co-creation on the cognitive and socio-affective outcomes of learning. Second, the effects of rubric co-creation in a more interactive setting of formative assessment, e.g., peer assessment, were understudied. Hence, the present study sets out from where they stopped to delve deeper into rubric co-creation from a standpoint of collaborative learning.

Rationale and Research Questions

The aim of the current study is to examine the effects of rubric co-creation on peer feedback in a setting of higher education. The study is grounded on the accounts of Carless and Boud (2018) that the quality of feedback information and opportunities for student involvement are two prerequisites for feedback success.

To attain such an aim, the present study is guided by the following logic. To begin with, the study conceptualizes learners as subjects of dual identity (as *assessors* and *assesseees*) who would garner learning gains from the interactive and dialogic processes involved in providing and using peer feedback (Gielen and De Wever 2015; Nicol, Thomson, and Breslin 2014). Next, rubric co-creation could elicit a deeper understanding of assessment criteria through negotiation and internalization during

developmental processes (Taylor, Kisby, and Reedy 2024). Consequently, the knowledge gains would impact the peer feedback processes: (1) from an assessor perspective, it facilitates the production of meaningful and actionable feedback; (2) from an assessee perspective, it allows feedback receivers to not only critically appreciate feedback messages but also participate willingly in post-reception discussion with the feedback provider, shall there be obscurities pending clarification. Finally, the peer interactions would result in a deepened understanding of the feedback comments and thereby influence actions to be taken (Zhao et al. 2023; Zhu and Carless 2018). Therefore, it is hypothesized that rubric co-creation would have positive effects on the quality, interactivity, and uptake of peer feedback. The following research questions are raised:

RQ1: What are the effects of rubric co-creation on the quality of peer feedback messages?

RQ2: What are the effects of rubric co-creation on the dialogic interactivity in peer feedback processes?

RQ3: What are the effects of rubric co-creation on the uptake of peer feedback messages?

Methodology

Participants and Context

The study took place in an English Language Studies program at a Chinese university.

The School of Foreign Languages has been implementing curriculum-embedded, student-centered, and learning-oriented feedback and assessment practices such as the formative use of evaluative instruments with students' involvement.

The participants of the study were second year undergraduate students from two intact classes of English Language study majors ($N=57$, $M_{age}=20.2$ years). The two classes were assigned to two interventional conditions: 30 in the rubric co-creation condition and 27 in the condition using the co-created rubric.

A pre-project English writing test following the requirement of Test for English Majors band 4 (TEM-4), a nation-level English proficiency test for English language majors in China, was administered. There were no statistically significant differences in TEM-4 writing measurement between the two groups ($F(1, 55) = 3.31$, $p = .074$; $\eta^2_p = 0.06$). Moreover, 18 students from the rubric co-creation group (60%) and 14 students from the rubric utilization group (51.9%) had previous experience using teacher-devised rubrics; none reported prior experience with rubric development.

The School of Foreign Languages approved the study to collect relevant data without disclosing participant identity and privacy in 2023. The students were informed of the policy of voluntary participation and their right to withdraw from the project. Written informed consent forms were obtained from all participants.

Conditions

The study followed the design of experimental conditions by Fraile, Panadero, and Pardo (2017): the rubric co-creation group (RC) and the rubric utilization group (RU) used the same peer feedback rubric co-created by the former. Since the two conditions shared the same teacher, instructional structure, and assessment tasks, the co-created rubric could be seamlessly used by the RU members.

Following the suggested procedures to co-create rubric (Kilgour et al. 2020; Jönsson and Panadero 2017), the rubrics were created through iterative discussions across and within student panel groups with the support of the teacher. The co-creation procedures are illustrated in **Figure 1**.

1. [Pre-creation session] Providing necessary training for rubric development; informing the learning objectives and task specifications; forming student rubric co-creation panels (in groups of 4 students)
2. [Pre-creation session] Gathering existing rubrics (from literature and teacher-devised) as exemplars.
3. [Pre-creation session] Meeting to determine the structure and timeframes for the co-creation activities.
4. [Session 1 of co-creation] Reaching consensus in assessment brief and feedback criteria for the rubric.
5. [Session 1] The students developing quality descriptors; teacher providing scaffolding and support; forming a draft version of the rubric.
6. [Session 2] Linking the rubric to task requirement, learning objectives and/or other accreditation standards (i.e., curriculum standards or competence frameworks); forming an improved version of the rubric.
7. [Session 2] Refining and finalizing the co-created rubric.

Figure 1. Rubric co-creation procedures

For the writing tasks in the project, the RC group members co-created three task-level feedback rubrics. See **Appendix A** for a sample rubric.

Procedures

The study was embedded in the course ‘English Writing II’, which lasted for 14 weeks (**Figure 2**). The study employed a quasi-experimental approach, utilizing a 2 x 3 mixed factorial design. The within-groups factor was the three writing tasks (T1 to T3) and the between-groups factor was the experimental conditions (rubric co-creation vs. rubric utilization). A few measures have been taken to ensure student engagement: (1) peer feedback was conducted in the form of written dialogic peer feedback on web-based collaborative platforms (Wood 2022; Zhao et al. 2023); (2) students were offered feedback templates as procedural facilitators reducing potential difficulties in constructing meaningful and actionable feedback messages (Gielen and De Wever 2015); (3) an assessment culture encouraging brave and agentic involvement in feedback provision and dialogs has been sustainably implemented in the project; and (4) remarkable performance and involvement in the project were awarded bonus marks in the final grades of the course.

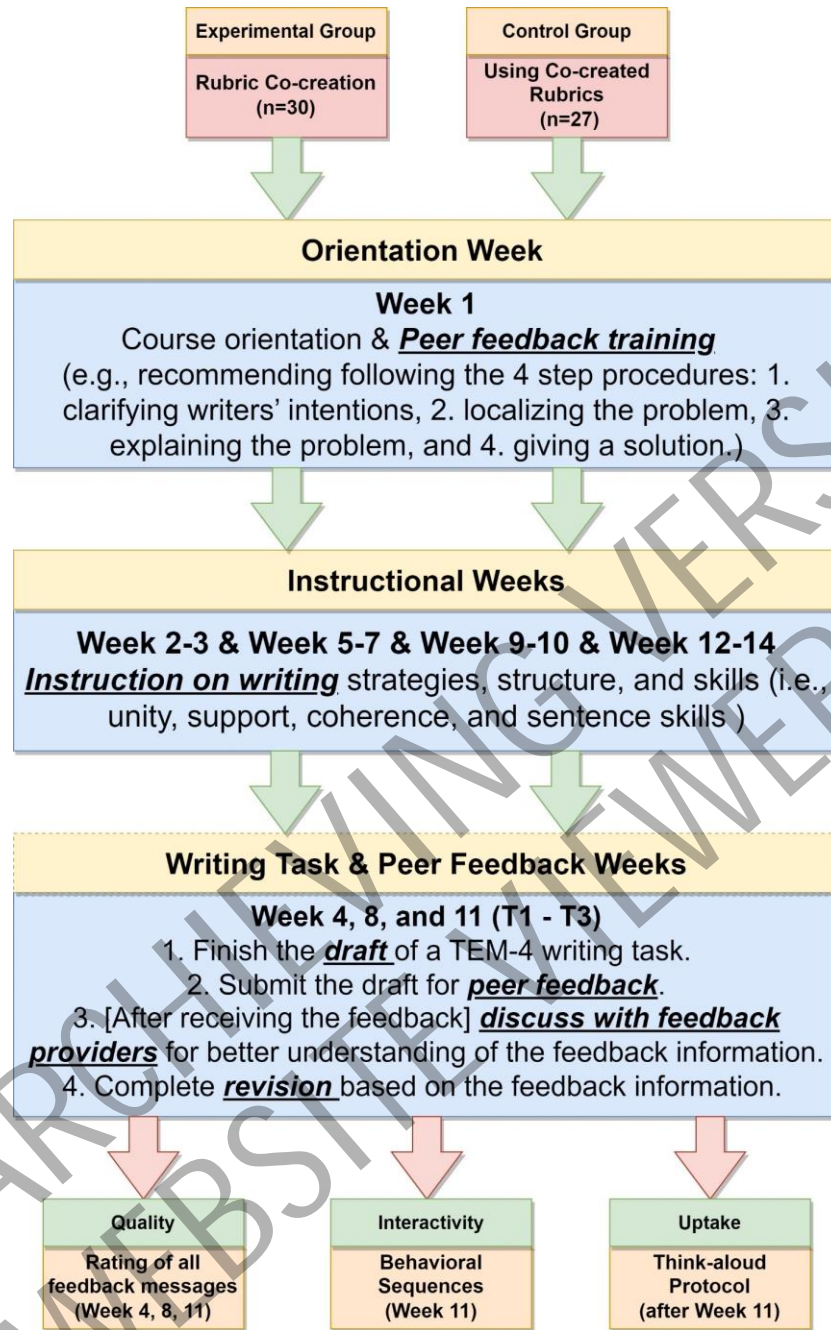


Figure 2. Experimental procedures

On weeks 4, 8, and 11, the students were required to complete three writing tasks with various topics. Corresponding rubrics were co-created by the RC group members and provided to the RU group members one week before the assessment weeks (i.e., on weeks 3, 7, and 10). Like the pre-project evaluation of writing

proficiency, the three tasks were designed in accordance with the requirements of TEM-4. Hence, the teacher provided a passage based on which the students were required to summarize and critically comment. According to the calculation by ARTE (<https://linguisticanalysistools.org/arte.html>), difficulties of the materials were equivalent to US reading grade level 13-16 (mean Flesch Reading Ease readability score was 43.8), on par with TEM-4 official tasks and suitable for second year English majors. The students completed draft writing on Shimo Docs, an online collaborative office suite.

Students were assigned to peer groups of 3 members. When the draft writing was submitted online, the participants reviewed and offered rubric-referenced written feedback for the writing artifacts from their peer learners against the criteria stipulated in the co-created rubrics. Hence, each participant provided and received two feedback messages for each of the three writing tasks. Cross-group exchange and discussion of feedback messages and subsequent revisions were prohibited throughout the study. Based on the received feedback, the students were encouraged to reciprocally discuss with the feedback providers to clarify the feedback information and finalize revision actions. After writing and feedback weeks, the drafts and revised writings were submitted to the teacher as learning artifacts of the course.

Measure

To measure feedback quality, a scoring rubric of the quality of peer feedback

messages by Gielen and De Wever (2015) was applied. The scoring rubric encompassed three major subcategories: use of criteria, feedback contents, and writing style. All 342 peer feedback messages for the three tasks were collected and rated by 5 trained co-researchers. The raters independently rated 50 feedback messages from T1 with a good inter-rater reliability ($ICC = .85$). They split the remaining rating duties.

To measure the dialogic interactivity in the peer feedback processes, behavioral data of feedback provision and discussions were retrieved from Shimo Docs and coded. The coding scheme was developed on top of the three-phase framework of dialogical peer feedback proposed by Zhao et al. (2023). A formative approach was adopted in its development with the researcher observing students' interactivity and enriching the categories in the coding scheme in week 4. The coding scheme (**Appendix B**) included 5 categories: initial feedback (IF); feedback-on-feedback (FF); feedback reception without inquiry (NI); re-feedback (RF); and not responding to feedback-on-feedback (NR). Compared to the framework by Zhao et al. (2023), two passive categories (i.e., feedback reception without inquiry and ignoring feedback-on-feedback) were added to capture students' behaviors in fruitless episodes or superficial attempts of feedback dialogs that emerged in T1. The participants' behaviors in T3 were collected and analyzed in the main study. Three co-researchers coded the data with a good inter-rater reliability (Fleiss's $Kappa = .78$).

Finally, students' responses to and actions upon peer feedback messages were measured as feedback uptake. Based on extant literature (van der Pol, Mercer, and Volman 2019; Barron 2003), uptake was coded into three categories: (1) ignore, (2) copy/repeat, and (3) apply. After the revision of writing task T3, all participants explained their actions taken upon all the received feedback points in an audio-recorded think-aloud session. After data cleaning and audio trimming, 312-minute audio data were transcribed for coding. The coding process included: (1) extracting focal sentences from students' verbalizations; (2) forming codes based on the focal sentences; and (3) categorizing the codes into the three major categories. Two co-researchers coded the data with a good inter-rater reliability (Cohen's Kappa = .95).

Analysis

The study analyzed three types of data: (1) quality rating of peer feedback messages; (2) the behaviors of dialogic interactions between peers; and (3) feedback uptake.

First, the researcher conducted a 2x3 split-plot ANOVA in R statistical software with feedback quality scores as dependent variables while the intervention groups (between-subjects) and writing tasks (within-subjects) as independent variables. Greenhouse-Geisser correction was applied to data violating the assumption of sphericity. Bonferroni correction was utilized to control the probability of a Type I error from multiple hypothesis tests between different time points.

Second, to examine the behavioral patterns of peer feedback interactions, lag sequential analysis (LSA) was performed using GSEQ 5.1. The software calculated adjusted residuals from a transitional probability matrix based on the coded behavior sequences (Pohl, Wallner, and Kriglstein 2016). The significance of behavioral transitions was determined by the Z-score of the adjusted residuals (significant if $Z > 1.96$). Behavioral transitions were visualized to present the interactional patterns within the dialogic peer feedback processes.

Finally, to compare feedback uptake between the two groups, Wilcoxon rank-sum tests instead of independent t-tests were applied since the violation of normality assumption requirement (Shapiro-Wilk test, $p < .05$).

Results

Feedback Quality

Split-plot ANOVA tests were run to analyze the effects of the rubric co-creation on the overall feedback quality scores and the three constituent subcategories (i.e., use of criteria, feedback contents, and writing style).

Regarding the overall feedback quality scores, the data showed that the experimental conditions ($F(1,55)=44.08, p<.001, \eta^2_p=.445$) and writing tasks ($F(1.81,99.51)=29.36, p<.001, \eta^2_p=.348$) had significant effects. Nevertheless, there was no significant interaction between the experimental conditions and writing tasks ($F(1.81,99.51)=1.86, p=.164, \eta^2_p=.033$). Post-hoc pairwise comparison revealed that

the rubric co-creation group (RC) outperformed the rubric utilization group (RU) in T2 ($t(55)=5.541, p<.001, d=0.868$) and T3 ($t(55)=5.380, p<.001, d=0.886$) while no significant difference was found in T1 ($t(55)=1.816, p=.075, d=.430$).

Regarding the use of criteria, experimental conditions ($F(1,55)=41.40, p<.001, \eta^2_p=.429$) and writing tasks ($F(1.54,84.56)=49.91, p<.001, \eta^2_p=.476$) had significant effects. However, significant interaction between the experimental conditions and writing tasks was not found ($F(1.54,84.56)=1.12, p=.318, \eta^2_p=.020$). Post-hoc pairwise comparison revealed that the RC group significantly outperformed the RU group in T1 ($t(55)=3.57, p<.001, d=0.747$), T2 ($t(55)=4.15, p<.001, d=0.945$), and T3 ($t(55)=4.29, p<.001, d=0.544$).

In terms of feedback contents, the experimental conditions had significant effects ($F(1,55)=10.70, p=.002, \eta^2_p=.163$) whereas writing tasks had no significant effects ($F(1.49,82.16)=0.04, p=.927, \eta^2_p=.001$). However, there was a significant interaction between the experimental conditions and writing tasks ($F(1.49,82.16)=6.40, p=.006, \eta^2_p=.104$). Simple main effect analysis and post-hoc pairwise comparison showed that the RC group outperformed the RU group in both T2 ($t(55)=4.08, p<.001, d=0.467$) and T3 ($t(55)=4.942, p<.001, d=0.796$).

Concerning feedback writing style, there were no significant effects of experimental conditions ($F(1,55)=0.61, p=.439, \eta^2_p=.011$) while writing tasks had significant effects ($F(1.53,84.42)=11.51, p<.001, \eta^2_p=.173$). Nonetheless, no significant interaction between the experimental conditions and writing tasks was

identified ($F(1.53, 84.42) = 2.44, p = .106, \eta^2_p = .042$). Post-hoc pairwise comparisons revealed that the RC group significantly outperformed the RU group in T1 ($t(55) = 2.117, p = .039, d = 0.252$) whereas no significant differences were found in T2 and T3 ($ps > .05$).

Feedback Interaction Patterns

The results of LSA are shown in **Appendix C**. Each row of the table represents an initial behavior while each column shows a subsequent behavior. For example, the IF→FF sequence of the RC group members displayed the behavioral transition of ‘after the reception of *initial feedback*, the receivers produced *feedback-on-feedback*’. Such behavioral transition was statistically significant ($Z_{IF \rightarrow FF} = 9.699$), indicating a strong trend to formulate back evaluation of the received feedback.

The behavioral patterns reflected by the LSA results are visualized in **Figure 3**. The RC group practiced dialogic peer feedback featuring a three-phase process (indicated by the IF→FF→RF route) and a bi-directional communication channel between feedback providers and receivers regarding feedback messages in need of explanation, clarification, and/or re-generation (indicated by IF↔RF). Comparatively, interactive peer feedback was not equally enacted in the RU group, particularly for the lack of (indicated by the weak RF→FF transition) and the tendency to terminate (indicated by the strong IF→NI and FF→NR transitions) dialogic communication.

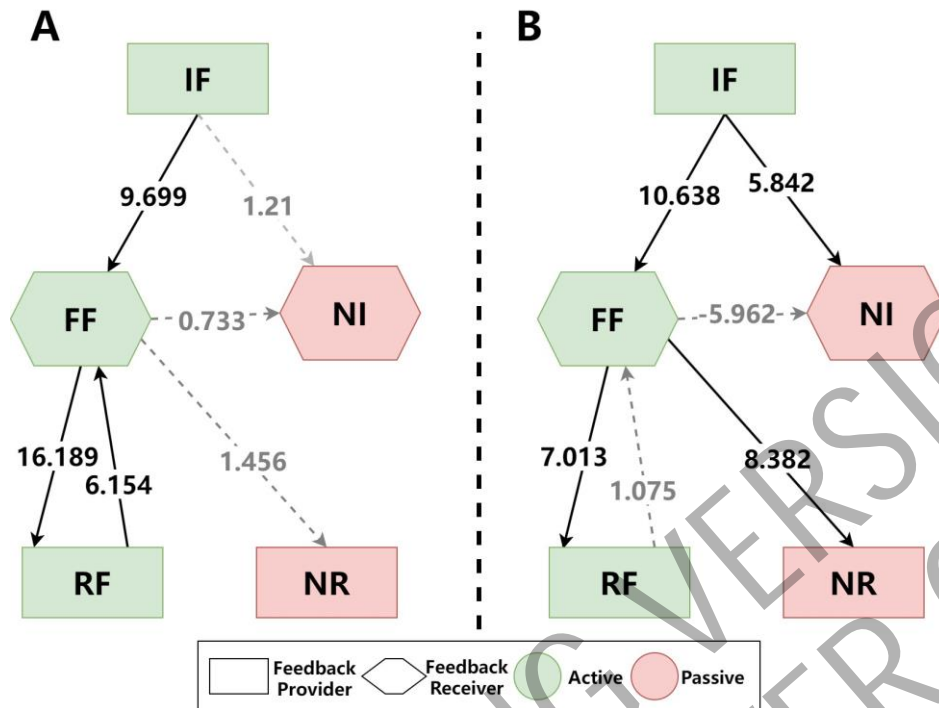


Figure 3. Behavioral patterns of peer feedback interactions.

Note: A: RC group; B: RU group. Spotted lines indicate insignificant behavioral transitions. IF: Initial Feedback; FF: Feedback-on-feedback; NI: Feedback Reception without Inquiry; RF: Re-feedback; NR: Not Responding to Feedback-on-feedback.

Feedback Uptake

The results from the Wilcoxon rank-sum tests are shown in **Table 1**. Regarding feedback uptake, the RC and RU groups had statistically significant differences in the categories of apply while there were no statistically significant differences in the uptake frequencies under the categories of ignore and copy/repeat. Compared against the effect size interpretation guidance by Gignac and Szodorai (2016), the effect sizes of the differences in the ignore and copy/repeat categories were small whereas that in the category of apply was large.

Table 1. Descriptive statistics and Wilcoxon rank-sum tests results of feedback uptake.

Uptake	RC		RU		Wilcoxon rank-sum tests			
	Mean	SD	Mean	SD	W	z	p	r
Ignore	1.9	1.3	2.8	2.5	321	-1.343	.183	.178
Copy/ Repeat	2.7	1.1	3.2	1.4	316	-1.422	.155	.188
Apply	6.4	2.3	2.7	2.4	678	4.363	<.001***	.578

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Discussion and Conclusion

The primary aim of the present study was to explore the quality, interactivity, and uptake of peer feedback, recognizing that students' involvement in co-creating and using rubrics may differ (Fraile, Panadero, and Pardo 2017). The topic is of central importance for formative assessment practices because issues pertinent to the cognitive and sociocultural aspects of peer feedback (Carless and Boud 2018) should be tackled in contemporary higher education (Gielen and De Wever 2015; Hattie and Timperley 2007). Subsequently, the results and corresponding implications were discussed against existing theoretical and empirical evidence.

Peer Feedback Quality

The study found an overall positive impact of rubric co-creation on the quality of peer feedback. As indicated by the split-plot ANOVA results, the intervention (i.e., rubric co-creation) significantly enhanced students' use of assessment criteria and the provision of effective feedback messages to peer learners (Panadero and Romero

2014). Such effects were in tandem with the viewpoints of Kilgour et al. (2020) that rubric co-creation could not only broaden students' understanding of the criteria and the overall assessment process, but also make the feedback process fairer and messages more objective. However, the results were inconsistent with Fraile, Panadero, and Pardo (2017) who argued that rubric co-creation has insignificant effects on students' perceptions of rubric use. On the one hand, the disparities could be explained by methodological differences; they measured students' perceptions through a self-reported questionnaire whereas the present study used teacher-rated quality scores. Noticeably, they inferred that the effects of rubric co-creation might need longer enactments to significantly change students' perceptions of rubric use. Juxtaposing the two studies, it is probably the case that students exhibited a delayed recognition of the impact of co-creation on their understanding and usage of assessment criteria. On the other hand, the differences could be attributed to the use of dialogic peer feedback in the present study. The significant improvement in criteria use and feedback contents could be seen as the outcomes of the iterative negotiation and meaning-making processes and the collaborative and interactive interpersonal relations in feedback practices (Er, Dimitriadis, and Gašević 2021; Zhao et al. 2023).

However, the intervention did not have a significant effect on writing styles of feedback, i.e., structure, use of keywords, and the first-person stance. To a degree, the insignificant effect justified the complexity of feedback practice and literacy (Carless and Boud 2018). According to Han and Xu (2020), feedback literate students must

possess an array of knowledge and capabilities ranging from language abilities to emotion regulation capabilities. Through such a lens, the results indicated that rubric co-creation was relatively ineffective in promoting students' metacognitive capability (e.g., the ability to monitor and reflect upon the structure of the feedback message before offering it to peer assesseses), feedback-related epistemological knowledge (e.g., the understanding of the key characteristics of effective and high quality feedback) and social-affective readiness (e.g., the abilities to establish both self-confidence and camaraderie), necessities for the production of feedback messages that match 'receiver preference' in styles (Prins, Sluijsmans, and Kirschner 2006, 292). Therefore, additional support beyond criteria co-construction is still needed for successful peer feedback practices. Previous literature has thrown light on this aspect, e.g., offering cognitive and socio-affective scaffolding from teachers (Xu and Carless 2017), providing procedural facilitators for students to generate well-structured feedback messages (Gielen and De Wever 2015). In sum, the present study added new insights to the compendium of instructional, procedural, and material supports available to enhance peer feedback quality (Bürgermeister, Glogger-Frey, and Saalbach 2021).

Patterns of Peer Feedback Interactivity

LSA of behavioral data identified two major findings. First, feedback receivers in the RC group demonstrated a higher level of agency in feedback, i.e., having more

constructive involvement in peer feedback discussion seeking assistance, clarification, and explanation through feedback-on-feedback or back-evaluation (van der Pol et al. 2008). Second, feedback providers in the RC group also showed a higher level of agency in feedback by actively participating in the feedback discussions initiated by feedback receivers (van den Berg, Admiraal, and Pilot 2006; Zhao et al. 2023). The above findings could be interpreted as a bipartite classification of student agency based on the dual identity of learners in dialogic peer feedback (Gielen and De Wever 2015). The results showcased that the intervention of rubric co-creation had significant effects on promoting agency while interacting as both feedback providers and receivers. In previous literature, agency in feedback has been predominantly conceptualized as a part of self-regulation in the context of self-assessment (Fraile, Panadero, and Pardo 2017; Panadero, Lipnevich, and Broadbent 2019). Alternatively, the present study indicated that student agency in peer feedback is realized through collaborative efforts in constructing and seeking meaningful feedback as well as specifying plans for further actions (Zhang et al. 2023; Zhao et al. 2023). On the one hand, such insights expanded our understanding of the effects of rubric co-creation on feedback agency; on the other hand, they conformed with recent viewpoints that peer feedback is a venue for co-regulation and authorial agency (Nieminen et al. 2022; Wood 2022).

Taken together, the LSA results further validated the three-step procedural framework of dialogic peer feedback proposed by Zhao et al. (2023), who argued that

effective dialogic feedback processes should encompass constructive interactivity in the form of feedback-on-feedback and re-feedback. The bidirectional communication between peer learners in the RC group echoed existing literature articulating the significance of idea exchange and meaning co-construction in peer feedback processes (Kilgour et al. 2020; Zhu and Carless 2018). Concurrently, the LSA results of the RU group revealed that peer feedback dialogs are not always effective and productive with learners avoiding or incapable of active involvement with peer interactions. Such a phenomenon could be explained by the accounts of Wu and Schunn (2023) on the multiple forms of engagement with peer feedback, i.e., a continuum inclusive of passive, active, and constructive engagement.

Peer Feedback Uptake

The research results revealed that the intervention had a significant effect on promoting critical usage of peer feedback messages. This has delivered a new understanding regarding the effectiveness of rubric co-creation on revision behaviors and feedback uptake since such effects in the context of peer feedback had not been documented in previous literature (e.g., Fraile, Panadero, and Pardo 2017).

Noteworthy, the finding partially echoed Fraile, Panadero, and Pardo (2017, 74) where the better performance of the experimental group was attributed to the usage of thinking-aloud protocol in the writing task for enhancing ‘exposure to the use of the internalized criteria’. Following the same line of reasoning, the critical feedback

uptake of the experimental group in the present study could be attributed to both peer feedback and rubric co-creation, which jointly augmented students' exposure to and abilities to apply the criteria (Kilgour et al. 2020; Panadero and Romero 2014).

However, feedback uptake is subject to an array of individual, contextual, and instructional factors (Kerman, Banihashem, and Noroozi 2023). The findings were in line with insights regarding the relationship between these factors and uptake enhancement. For example, Wu and Schunn (2020) had argued that students receiving high quality peer feedback are more likely to actively uptake and incorporate the feedback messages into writing artifacts. Similarly, Zhang et al. (2021) examined the relationship between optimal behavioral sequences and better uptake of peer feedback. In the present study, the intervention has facilitated the two prerequisites for improved feedback uptake. Hence, the direct and indirect impacts of rubric co-creation on uptake improvement should be holistically interpreted and comprehended.

Implications

The findings have several implications for educators poised to incorporate the co-construction of assessment criteria into formative assessment practices. The research identified the potential of rubric co-creation to support successful peer feedback. To date, rubric co-creation in a collaborative learning setting remains under-implemented (Taylor, Kisby, and Reedy 2024). Therefore, the application of rubric co-creation as a pedagogical intervention should be encouraged in various domains of education.

However, the findings also revealed that rubric co-creation was not an all-round game-changer for peer feedback challenges. Thus, scaffolding and support from a multitude of channels and perspectives should be considered to continuously augment students' ownership of and capabilities in assessment practices.

Furthermore, the study revealed that rubric co-creation could elevate learner interactivity and involvement in dialogic feedback processes. Therefore, educators are advised to transform from an old paradigm of feedback practice to a new one, where the role of students is recast as active feedback seekers and negotiators (Carless 2015; Winstone and Carless 2019). However, it should be noted that the researcher applied several stimuli for engagement and agency in feedback (see **Procedures**). Educators should further explore effective approaches to promote dialogic peer feedback since the above-mentioned stimuli might be either time-consuming or unsustainable for longer-duration implementation.

Limitations and Future Directions

Being a pioneering effort to examine the effects of rubric co-creation on peer feedback, the study inevitably carries some limitations. First, the present study shed no light on the contents of feedback dialogs between peer learners. Identifying behavioral patterns might not suffice for an in-depth understanding of the interactive processes within peer feedback. In follow-up research, researchers could use qualitative approaches to explore what and how peer learners discuss after feedback

reception, and their perceptions thereof. Second, the study only examined the quality, interactivity, and uptake of peer feedback. Considering the complexity of peer feedback, such limitation is understandable since it is impossible to thoroughly study the nature of peer feedback in a single paper. However, existing literature has underlined other features of peer feedback worth studying, e.g., problem identification, localization, scope, praise, and mitigating language (Nelson and Schunn 2009). Subsequent studies are encouraged to delve deeper into these features of peer feedback in a setting that practices rubric co-creation. Finally, a single rubric co-creation approach was applied in the study. Hence, we are not aware of the differences between various artifacts of criteria co-creation. Successive studies could enrich available strategies to suit domain-specific and context-dependent needs and study their impacts comparatively.

Acknowledgements

The author thanks the handling editor and the three anonymous reviewers who have provided constructive feedback for the quality improvement of the paper.

Notes on contributor

Da Yan is a senior lecturer at Xinyang Agriculture and Forestry University, China. His research interests include formative assessment, use of rubric in education, and GenAI-enhanced instructional feedback.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This research was funded by the Young Researcher Program of Xinyang Agriculture and Forestry University [grant number QN2022049].

References

- Ashton, S., and R. S. Davies. 2015. "Using Scaffolded Rubrics to Improve Peer Assessment in a MOOC Writing Course." *Distance Education* 36 (3): 312–334. doi:10.1080/01587919.2015.1081733.
- Bacchus, R., E. Colvin, E. B. Knight, and L. Ritter. 2020. "When Rubrics Aren't Enough: Exploring Exemplars and Student Rubric Co-Construction." *Journal of Curriculum and Pedagogy* 17 (1): 48–61. doi:10.1080/15505170.2019.1627617.
- Barron, B. 2003. "When Smart Groups Fail." *Journal of The Learning Sciences* 12 (3): 307–359. doi:10.1207/S15327809JLS1203_1.
- Berg, I. van den, W. Admiraal, and A. Pilot. 2006. "Peer assessment in university teaching: evaluating seven course designs." *Assessment & Evaluation in Higher Education* 31 (1): 19–36. doi:10.1080/02602930500262346.
- Black, P., and D. Wiliam. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability* 21 (1): 5–31. doi:10.1007/s11092-008-9068-5.
- 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. doi:10.1080/02602938.2012.691462.
- Brookhart, S. M. 2018. "Appropriate Criteria: Key to Effective Rubrics." *Frontiers in Education* 3 (22). doi:10.3389/educ.2018.00022.
- Brookhart, S. M., and F. Chen. 2015. "The Quality and Effectiveness of Descriptive Rubrics." *Educational Review* 67 (3): 343–368. doi:10.1080/00131911.2014.929565.
- Bürgermeister, A., I. Glogger-Frey, and H. Saalbach. 2021. "Supporting Peer Feedback on Learning Strategies: Effects on Self-Efficacy and Feedback

- Quality." *Psychology Learning & Teaching* 20 (3): 383–404.
doi:10.1177/14757257211016604.
- Carless, D. 2015. *Excellence in University Assessment: Learning from Award-Winning Practice*. London New York: Routledge, Taylor & Francis Group.
- Carless, D., and D. Boud. 2018. "The Development of Student Feedback Literacy: Enabling Uptake of Feedback." *Assessment & Evaluation in Higher Education* 43 (8): 1315–1325. doi:10.1080/02602938.2018.1463354.
- Er, E., Y. Dimitriadis, and D. Gašević. 2021. "A Collaborative Learning Approach to Dialogic Peer Feedback: A Theoretical Framework." *Assessment & Evaluation in Higher Education* 46 (4): 586–600.
doi:10.1080/02602938.2020.1786497.
- Fraile, J., E. Panadero, and R. Pardo. 2017. "Co-Creating Rubrics: The Effects on Self-Regulated Learning, Self-Efficacy and Performance of Establishing Assessment Criteria with Students." *Studies in Educational Evaluation* 53: 69–76. doi:10.1016/j.stueduc.2017.03.003.
- Gielen, M., and B. De Wever. 2015. "Structuring the Peer Assessment Process: A Multilevel Approach for the Impact on Product Improvement and Peer Feedback Quality." *Journal of Computer Assisted Learning* 31 (5): 435–449.
doi:10.1111/jcal.12096.
- Gignac, G. E., and E. T. Szodorai. 2016. "Effect Size Guidelines for Individual Differences Researchers." *Personality and Individual Differences* 102: 74–78.
doi:10.1016/j.paid.2016.06.069.
- Gong, H., and D. Yan. 2023. "The Impact of Danmaku-Based and Synchronous Peer Feedback on L2 Oral Performance: A Mixed-Method Investigation." *PLOS ONE* 18 (4): e0284843. doi:10.1371/journal.pone.0284843.
- Hamer, J., H. Purchase, A. Luxton-Reilly, and P. Denny. 2015. "A Comparison of Peer and Tutor Feedback." *Assessment & Evaluation in Higher Education* 40 (1): 151–164. doi:10.1080/02602938.2014.893418.
- Han, Y., and Y. Xu. 2020. "The Development of Student Feedback Literacy: The Influences of Teacher Feedback on Peer Feedback." *Assessment & Evaluation in Higher Education* 45 (5): 680–696. doi:10.1080/02602938.2019.1689545.
- Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.
- Huisman, B., N. Saab, J. van Driel, and P. van den 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. doi:10.1080/02602938.2018.1424318.
- Jönsson, A., and E. Panadero. 2017. "The Use and Design of Rubrics to Support Assessment for Learning." In *Scaling up Assessment for Learning in Higher Education*, edited by David Carless, Susan M. Bridges, Cecilia Ka Yuk Chan, and Rick Glofcheski, 99–111. The Enabling Power of Assessment. Singapore: Springer. doi:10.1007/978-981-10-3045-1_7.

- Kerman, N. T., S. K. Banihashem, and O. Noroozi. 2023. "The Relationship among Students' Attitude towards Peer Feedback, Peer Feedback Performance, and Uptake." In *The Power of Peer Learning: Fostering Students' Learning Processes and Outcomes*, edited by Omid Noroozi and Bram De Wever, 347–371. Social Interaction in Learning and Development. Cham: Springer International Publishing. doi:10.1007/978-3-031-29411-2_16.
- Kilgour, P., M. Northcote, A. Williams, and A. Kilgour. 2020. "A Plan for the Co-Construction and Collaborative Use of Rubrics for Student Learning." *Assessment & Evaluation in Higher Education* 45 (1): 140–153. doi:10.1080/02602938.2019.1614523.
- Man, D., B. Kong, and M. H. Chau. 2022. "Developing Student Feedback Literacy through Peer Review Training." *RELC Journal*, 336882221078380. doi:10.1177/00336882221078380.
- 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. doi:10.1007/s11251-008-9053-x.
- Nicol, D. 2010. "From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education." *Assessment & Evaluation in Higher Education* 35 (5): 501–517. doi:10.1080/02602931003786559.
- Nicol, D., A. Thomson, and C. Breslin. 2014. "Rethinking Feedback Practices in Higher Education: A Peer Review Perspective." *Assessment & Evaluation in Higher Education* 39 (1): 102–122. doi:10.1080/02602938.2013.795518.
- Nieminen, J. H., J. Tai, D. Boud, and M. Henderson. 2022. "Student Agency in Feedback: Beyond the Individual." *Assessment & Evaluation in Higher Education* 47 (1): 95–108. doi:10.1080/02602938.2021.1887080.
- Panadero, E., and A. Jonsson. 2013. "The use of scoring rubrics for formative assessment purposes revisited: A review" *Educational Research Review* 9: 129–144. doi: 10.1016/j.edurev.2013.01.002.
- Panadero, E., and A. Jonsson. 2020. "A Critical Review of the Arguments against the Use of Rubrics." *Educational Research Review* 30: 100329. doi:10.1016/j.edurev.2020.100329.
- Panadero, E., A. Jonsson, and M. Alqassab. 2018. "Providing Formative Peer Feedback: What Do We Know?" In *The Cambridge Handbook of Instructional Feedback*, edited by Anastasiya A. Lipnevich and Jeffrey K. Smith, 409–431. Cambridge Handbooks in Psychology. Cambridge: Cambridge University Press. doi:10.1017/9781316832134.020.
- Panadero, E., A. Lipnevich, and J. Broadbent. 2019. "Turning Self-Assessment into Self-Feedback." In *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners*, edited by Michael Henderson, Rola Ajjawi, David Boud, and Elizabeth Molloy, 147–163. Cham: Springer International Publishing. doi:10.1007/978-3-030-25112-3_9.

- Panadero, E., and M. Romero. 2014. "To Rubric or Not to Rubric? The Effects of Self-Assessment on Self-Regulation, Performance and Self-Efficacy." *Assessment in Education: Principles, Policy & Practice* 21 (2): 133–148. doi:10.1080/0969594X.2013.877872.
- Pohl, M., G. Wallner, and S. Kriglstein. 2016. "Using Lag-Sequential Analysis for Understanding Interaction Sequences in Visualizations." *International Journal of Human-Computer Studies* 96: 54–66. doi:10.1016/j.ijhcs.2016.07.006.
- Pol, J. van der, B. A. M. van den Berg, W. F. Admiraal, and P. R. J. Simons. 2008. "The Nature, Reception, and Use of Online Peer Feedback in Higher Education." *Computers & Education* 51 (4): 1804–1817. doi:10.1016/j.compedu.2008.06.001.
- Pol, J. van der, N. Mercer, and M. Volman. 2019. "Scaffolding Student Understanding in Small-Group Work: Students' Uptake of Teacher Support in Subsequent Small-Group Interaction." *Journal of The Learning Sciences* 28 (2): 206–239. doi:10.1080/10508406.2018.1522258.
- Popham, W. J. 1997. "What's Wrong—and What's Right—with Rubrics." *Educational Leadership* 55 (2): 72–75.
- Prins, F. J., D. M. A. Sluijsmans, and P. A. Kirschner. 2006. "Feedback for General Practitioners in Training: Quality, Styles, and Preferences." *Advances in Health Sciences Education* 11 (3): 289–303. doi:10.1007/s10459-005-3250-z.
- Reddy, Y. M., and H. Andrade. 2010. "A Review of Rubric Use in Higher Education." *Assessment & Evaluation in Higher Education* 35 (4): 435–448. doi:10.1080/02602930902862859.
- Santana-Martel, J. S., and A. Perez-Garcías. 2022. "Students' Agency and Self-Regulated Skills through the Lenses of Assessment Co-Creation in Post-COVID-19 Online and Blended Settings: A Multi-Case Study." *Journal of Interactive Media in Education* 2022 (1): 1-17. doi:10.5334/jime.746.
- Taylor, B., F. Kisby, and A. Reedy. 2024. "Rubrics in Higher Education: An Exploration of Undergraduate Students' Understanding and Perspectives." *Assessment & Evaluation in Higher Education*, 1–11. doi:10.1080/02602938.2023.2299330.
- To, J. 2022. "Using Learner-Centred Feedback Design to Promote Students' Engagement with Feedback." *Higher Education Research & Development* 41 (4): 1309–1324. doi:10.1080/07294360.2021.1882403.
- Winstone, N., and D. Carless. 2019. *Designing Effective Feedback Processes in Higher Education: A Learning-Focused Approach*. 1st edition. London ; New York: Routledge.
- Wood, J. 2022. "Making Peer Feedback Work: The Contribution of Technology-Mediated Dialogic Peer Feedback to Feedback Uptake and Literacy." *Assessment & Evaluation in Higher Education* 47 (3): 327–346. doi:10.1080/02602938.2021.1914544.

- Wu, Y., and C. D. Schunn. 2020. "When Peers Agree, Do Students Listen? The Central Role of Feedback Quality and Feedback Frequency in Determining Uptake of Feedback." *Contemporary Educational Psychology* 62: 101897. doi:10.1016/j.cedpsych.2020.101897.
- Wu, Y., and C. D. Schunn. 2023. "Passive, Active, and Constructive Engagement with Peer Feedback: A Revised Model of Learning from Peer Feedback." *Contemporary Educational Psychology* 73: 102160. doi:10.1016/j.cedpsych.2023.102160.
- Xu, Y., and D. Carless. 2017. "'Only True Friends Could Be Cruelly Honest': Cognitive Scaffolding and Social-Affective Support in Teacher Feedback Literacy." *Assessment & Evaluation in Higher Education* 42 (7): 1082–1094. doi:10.1080/02602938.2016.1226759.
- Zhang, F., H.-T. Min, P. He, S. Chen, and S. Ren. 2023. "Understanding EFL Students' Feedback Literacy Development in Academic Writing: A Longitudinal Case Study." *Assessing Writing* 58: 100770. doi:10.1016/j.asw.2023.100770.
- Zhang, Y., Z. Pi, L. Chen, X. Zhang, and J. Yang. 2021. "Online Peer Assessment Improves Learners' Creativity: Not Only Learners' Roles as an Assessor or Assessee, but Also Their Behavioral Sequence Matter." *Thinking Skills and Creativity* 42: 100950. doi:10.1016/j.tsc.2021.100950.
- Zhao, K., J. Zhou, and P. Dawson. 2021. "Using Student-Instructor Co-Constructed Rubrics in Signature Assessment for Business Students: Benefits and Challenges." *Assessment in Education: Principles, Policy & Practice* 28 (2): 170–190. doi:10.1080/0969594X.2021.1908225.
- Zhao, Y., F. Zhang, C. D. Schunn, P. He, D. Li, and Y. Zhao. 2023. "Feedback, Feedback-on-Feedback and Re-Feedback: Effects of Written Dialogic Peer Feedback on English as a Foreign Language Writing." *Assessment & Evaluation in Higher Education*, 1–16. doi:10.1080/02602938.2023.2278017.
- Zhu, Q., and D. Carless. 2018. "Dialogue within Peer Feedback Processes: Clarification and Negotiation of Meaning." *Higher Education Research & Development* 37 (4): 883–897. doi:10.1080/07294360.2018.1446417.

Appendices

Appendix A: A sample co-created rubric (translated from Chinese)

Task Prompt: Do students majoring in liberal arts need to learn mathematics?

Dimensions	Competence*	Learning Objectives	Levels			
			Excellent	Good	Fair	Poor
Task Relevancy	Writing Strategy Competence Level 7-8	1. Identify key elements within a writing task, including purpose, audience, and specific requirements. 2. Analyze and evaluate the relevance of information, ideas, and arguments in relation to the given task.	Clearly identifies and thoroughly addresses all components of the prompt, showcasing a deep understanding of the relevance of mathematics to liberal arts. Arguments are consistently tied back to the task and prompt.	Effectively addresses the main aspects of the prompt, demonstrating a clear understanding of the task's relevance. Arguments are generally connected to the prompt with minor lapses.	Addresses the task but may have occasional lapses in relevance or understanding. Some arguments may lack a clear connection to the prompt.	Attempts to address the task but often strays off-topic or lacks clear relevance. Arguments are loosely connected or unclear.
Content	Overall Writing Competence Level 7. & Writing Argumentation, Narrative,	1. Develop the ability to generate well-supported ideas and arguments that contribute to a	Provides a comprehensive exploration of the topic, presenting well-supported arguments on why	Presents clear and relevant ideas, demonstrating a good understanding of the topic.	Offers adequate content with some development and relevant ideas, but may lack depth or originality.	Provides limited content with little development and minimal relevant ideas. Arguments

	Descriptive Competence Level 7	comprehensive exploration of a topic. 2.Synthesize information from various sources to strengthen the depth and richness of content.	mathematics is or isn't crucial for liberal arts students. Information is synthesized from multiple sources, enhancing the depth and richness of the content.	Arguments are well-developed and contribute to a thorough exploration of the prompt.	Arguments may need further support or clarification.	lack depth and sufficient support.
Organization	Overall Organization Competence Level 6-7	1.Structure writing logically with a clear introduction, body, and conclusion. 2.Use effective transitions to guide readers smoothly between ideas and paragraphs.	Exhibits an exceptionally well-organized structure with a clear introduction, body, and conclusion. Transitions guide the reader smoothly between ideas, enhancing coherence. Each paragraph contributes effectively to the overall flow of arguments.	Demonstrates a well-organized structure with a clear introduction, body, and conclusion. Transitions are effective, ensuring a logical flow of ideas. Each paragraph contributes to the overall organization, supporting the main arguments.	Generally organized but may have occasional issues with structure, transitions, or flow. The organization does not significantly hinder understanding. Some paragraphs may lack a clear contribution to the overall structure.	Organization is weak, making it difficult to follow the progression of ideas. Transitions may be abrupt or unclear. Contributions of paragraphs to the overall structure are unclear.
Language Use	Overall Language	1.Demonstrate a strong command of grammar,	Uses precise and varied language, demonstrating an	Employs clear and effective language, with few errors,	Mostly clear language with some errors that	Often unclear language with frequent errors that

	Competence Level 8 & Writing Argumentation, Narrative, Descriptive Competence Level 7	punctuation, and vocabulary appropriate for the audience and purpose. 2.Utilize different sentence structures and word choices to enhance clarity and engagement.	exemplary command of grammar, punctuation, and vocabulary. Sentences are well-constructed, enhancing clarity and engagement. Language choices effectively convey the writer's stance on the relevance of mathematics to liberal arts.	showcasing a strong command of grammar, punctuation, and vocabulary. Sentences are varied and contribute to overall effectiveness. Language choices generally convey the writer's stance on the relevance of mathematics.	may slightly impede understanding. The writing relies on a standard vocabulary and may lack variety. Sentence construction may need improvement. Language choices inconsistently convey the writer's stance.	significantly hinder understanding. The vocabulary is limited, and errors detract from the overall quality. Sentence construction needs significant improvement. Language choices do not effectively convey the writer's stance.
Creativity	Narrative Writing Competence Level 8 & Overall Organization Competence Level 9	1.Generate original ideas and perspectives to bring a unique and creative approach to writing. 2.Try different writing styles, literary devices, and narrative techniques.	Demonstrates exceptional creativity, showcasing original thinking, unique perspectives, and innovative ideas in presenting arguments on the role of mathematics in liberal arts. Creativity	Shows creativity with original ideas and a fresh perspective on the topic. Creativity enhances the overall quality and engages the reader in considering the role of mathematics in liberal arts.	Demonstrates some creativity, but may rely on common ideas or lack originality. Creativity contributes modestly to the overall quality and engagement with the topic.	Lacks creativity, relying on clichés or conventional ideas. Creativity is absent or minimal, hindering overall engagement with the topic.

			significantly enhances the overall quality and impact of the writing.			
Argumentation & Criticality	Critical Writing Competence Level 7-8	1. Present compelling arguments that are well-supported, logically structured, and demonstrate critical evaluation. 2. Engage effectively with counterarguments, showcasing a high level of critical thinking and demonstrating an understanding of the topic.	Presents compelling arguments that are well-supported, logically structured, and demonstrate a critical evaluation of relevant information. Engages with counterarguments effectively, showcasing a high level of critical thinking.	Presents solid arguments with some support and logical structure. Demonstrates a good understanding of the topic but may lack in-depth critical evaluation. Engagement with counterarguments is present but may be limited.	Presents arguments that are somewhat supported, with occasional lapses in logical structure. Critical evaluation of information may be limited. Engagement with counterarguments is minimal.	Presents weak or unsupported arguments with little logical structure. Critical evaluation is lacking, and engagement with counterarguments is absent or unclear.

Note: *: Competence points are in accordance with the *China Standards of English*, the national English competence framework of China. See <https://cse.neea.edu.cn/> for details (in Chinese).

In actual usage, the rubric is accompanied by writing exemplar selected and agreed-upon during the co-creation processes.

Appendix B: Coding scheme for interactive behaviors in dialogic peer feedback.

Code	Category	Definition	Example	Unit
IF	Initial Feedback	Feedback provider evaluate writing artifact of the peer learner and provide qualitative comments and suggestions about its strength and weakness without offering scoring or grading (Panadero, Jonsson, and Alqassab 2018).	<p><i>Here below are five comments for you to consider:</i></p> <p><i>1. Try to use a more formal style of language. For example, the “get the best from” in the second paragraph could be replaced by “make the best use of”. I suggest you check ludwig.guru if you need to check which one is better.</i></p> <p><i>... (omitted)...</i></p> <p><i>* Mandatory in the present study</i></p>	Per feedback message.
FF	Feedback-on-feedback	Feedback receivers inquired about “the problematic elements of the received peer feedback” to clarify vagueness in the feedback message, offering the original writing intentions, explain disagreement in understanding, or seeking ‘a more feasible revision suggestion’ (Zhao et al. 2023, 2).	<p><i>I partially understand your comments here. But I need to make it clear why I have originally written in this way.</i></p> <p><i>... (omitted)...</i></p> <p><i>Can you elaborate on this point? I have checked the following resources. It seems your criticism here are not accurate enough.</i></p> <p><i>... (omitted)...</i></p> <p><i>I would still belief it is hard to carry out, do you have an alternative solution to this issue?</i></p> <p><i>... (omitted)...</i></p>	Per attempt made by the peer assessee or assessor.
NI	Feedback Reception without Inquiry	Feedback receivers hold the feedback final (not necessarily accepting it and ready to apply it for revision) without intention to discuss with or seeking assistance from the feedback provider (Min, 2003).	<p><i>I have received the comments, and I will try to apply them.</i></p> <p><i>Thank you.</i></p> <p><i><Take direct actions based on the comment without any follow up responses on the platform></i></p>	Per attempt made by the peer assessee or assessor.

RF	Re-feedback	Feedback providers ‘clarify misunderstandings in previous comments, confirm or refine previous comments, or propose new comments’ (Zhao et al. 2023, 2).	<p><i>I think you have a misunderstanding in this regard. Please see this link for a source I found online which I think can help you.</i> ... (omitted)...</p> <p><i>I can add two more detailed point so that you can get a more in-depth understanding of the issue we are discussing now.</i> ... (omitted)...</p> <p><i>I think my original comments are erroneous, please see the following new ones I have re-written, I am very sorry for this.</i> ... (omitted)...</p>	
NR	Not Responding to Feedback-on-feedback	After offering the initial feedback, the feedback providers hold his duty fulfilled without trying to answer the calls from feedback receivers (van den Berg, Admiraal, and Pilot 2006).	<p><i>I think the comments themselves are already in detail, I can not help you more.</i> ... (omitted)...</p> <p><i><Ignoring the follow-up responses from the feedback receiver on the platform without any contact or action></i></p>	

References (excluding items in main reference list):

Min, H.T. 2003. “Why Peer Comments Fail?” *English Teaching & Learning* 27 (3): 85–103.

Appendix C: Results of LSA

Table Appendix-C1. Frequency transition and corresponding adjusted residual of the RC group.

		IF	FF	NI	RF	NR
IF	Freq.	0	178	21	15	0
	AR	0	9.699*	1.21	-10.723	-1.94
FF	Freq.	0	0	11	98	2
	AR	0	-15.829	0.733	16.189*	1.456
NI	Freq.	0	4	0	0	0
	AR	0	1.585	-0.604	-1.293	-0.178
RF	Freq.	0	56	0	0	1
	AR	0	6.154*	-2.459	-5.261	0.91
NR	Freq.	0	0	0	0	0
	AR	0	0	0	0	0

Note: * $Z > 1.96$; Freq.: Frequency; AR: Adjusted Residual for the behavioral transition.

Table Appendix-C2. Frequency transition and corresponding adjusted residual of the RU group.

		IF	FF	NI	RF	NR
IF	Freq.	0	61	52	10	0
	AR	0	10.638*	5.842*	-6.88	-8.271
FF	Freq.	0	0	24	85	79
	AR	0	-10.783	-5.962	7.013*	8.382*
NI	Freq.	0	0	0	0	0
	AR	0	0	0	0	0
RF	Freq.	0	1	1	0	0
	AR	0	1.075	0.837	-0.937	-0.824
NR	Freq.	0	0	0	0	0
	AR	0	0	0	0	0

Note: * $Z > 1.96$; Freq.: Frequency; AR: Adjusted Residual for the behavioral transition.