



Critical Reflection in Peer Feedback: Assessing the Evaluative Quality of Peer Comments in Collaborative Writing Tasks

Arif Putra Yogesi¹, La Sunra^{2*}, Muh. Tahir³

^{1,2,3} English Language Education Study Program, Graduate Program, Universitas
Negeri Makassar

Article Info	Abstract
Received: 2026 - 06- 01 Revised: 2026 06-18 Accepted: 2026 06-23	<i>Peer feedback is an integral part of collaborative writing; nevertheless, the evaluative depth of student comments is often limited. This study examines the quality of peer feedback among 15 participants in a mixed-age English language course where age differences and social hierarchy may influence evaluative interaction. This research employs Qualitative Content Analysis (QCA) and centers around two primary questions: to what extent do student peer comments reflect critical engagement versus technical corrections (RQ1), and what are the common patterns of evaluative reasoning (RQ2). Data were collected via a Peer Feedback Worksheet designed and assessed by an evaluation rubric, from Level 1 (Corrective) to Level 3 (Strategic). The results reveal that roughly 60% of student comments were superficial, largely addressing minor technical issues such as spelling and grammar. Qualitative patterns imply that social dynamics, notably age seniority and a need for social civility, greatly inhibit students from giving strategic, high-quality comments. The younger participants were uncomfortable evaluating their older peers, resulting in a "safety-first" approach to their reasoning. The study suggests that language proficiency is not sufficient for effective peer feedback and that educational interventions should target social-emotional feedback literacy to enhance deeper critical reflection in different EFL settings. The study also highlights the importance of social-emotional feedback literacy in improving peer-feedback pedagogy in EFL contexts.</i>
Keywords: Collaborative Writing, Critical Reflection, Evaluative Quality, Mixed-age Classroom, Peer feedback.	
DOI: 10.24256/ideasv14i1.10718	
Corresponding Author: Arif Putra Yogesi arifp5097@gmail.com English Language Education Study Program, Graduate Program, Universitas Negeri Makassar	

1. Introduction

In recent years, the approach to Second Language (L2) writing education has changed considerably from students working in solitude to more social,

collaborative contexts. Today, collaborative writing is a standard activity in many EFL classrooms, since it gives learners the opportunity to negotiate meaning and gain knowledge together. As Storch (2021) points out, the importance of these assignments lies not only in the final paper produced but also in the “*languageing*” or deep debates that take place as students write. Students observe diverse ways of using the language through these exchanges and this naturally helps them polish their grammar and their general writing style. However, the success of such interaction depends to a considerable extent on the quality of peer feedback amongst the students.

Peer feedback is a great tool in that it makes pupils more independent and less dependent on the teacher to make every correction. Hyland (2022) suggests that when a student gives comments, it is an opportunity for the student to think about their own linguistic standards and improve their own writing skills. In recent years, digital tools and messaging apps have made it easier than ever for students to swap opinions. However, earlier studies (Yu & Hu, 2021) show that the quality of peer evaluation is frequently reduced when students do not critically engage with their peer’s work.

One of the common problems in the L2 classroom is what we call “*surface-level*” feedback. Many students prefer to adopt low-risk evaluative strategies, offering generic comments like “*good job*” or merely changing small typos and spelling errors. They will shy away from the more difficult work of considering the real content or the logic of the argument. This indicates a lack of “*feedback literacy*” as the ability to both give and receive helpful, constructive criticism (Winstone et al., 2022). If remarks are too simplistic, pupils don’t stretch themselves intellectually, and the entire collaborative effort becomes an exercise in squandered opportunities for higher order thinking.

Critical reflection must be emphasised in order to fix this. Critical reflection in this study, is not merely a matter of recognising an error, it is the ability of the student to identify a particular problem, explain why it is a problem and propose a viable remedy. The quality of a student’s reply, Lee (2023) says, genuinely reflects the student’s understanding of the objective of the material they are reading. But without high-quality evaluation, the peer review typically becomes a polite exercise, rather than an opportunity for academic improvement.

In addition, this study is informed by Brown and Levinson’s Face Theory, which explains how individuals attempt to maintain interpersonal harmony during interaction. In peer-feedback activities, students may avoid direct criticism in order to protect their peers’ social “*face*,” particularly in mixed-age classrooms where age hierarchy and politeness are culturally significant. Consequently, students may prioritize politeness over critical engagement, which can reduce the depth of evaluative feedback.

Despite these challenges, previous studies also demonstrate that peer feedback can become highly effective when students receive sufficient scaffolding and feedback training. Rahimi (2021) explains that teacher guidance can significantly improve evaluative interaction by developing students’ confidence and metalinguistic awareness. Similarly, Min (2016) found that explicit peer-review training enables learners to provide more analytical and strategic feedback.

Although there is a growing body of literature on peer feedback in EFL

contexts, past research has mostly concentrated on the effectiveness of peer feedback in enhancing writing achievement or linguistic accuracy (Yu & Hu, 2021; Liu & Hansen, 2022). Little emphasis has been paid to the evaluative quality of peer comments, particularly in terms of how students develop their rationale in giving feedback. Moreover, previous research mostly explores peer feedback within homogenous classroom contexts, with less consideration of how social elements like age disparities and interpersonal relationships influence students' evaluative behaviours.

Thus, there is still a need for research that analyses not only the extent of evaluative engagement in peer feedback but also the underlying patterns of reasoning in socially varied learning environments.

This study examined the quality of peer comments among participants in a private English language course. The students are from diverse age groups and experiences, so it provides a fresh lens on whether their comments are perceived as serious thought or just basic technical changes. The researcher aims to uncover trends in their comments that can enable us as instructors to better facilitate critical reflection in a varied classroom context. This study contributes to the growing body of research on peer feedback by emphasising the significance of social-emotional aspects on the quality of evaluation, particularly in mixed-age EFL environments. There are 2 research questions, to what extent do student peer comments reflect critical engagement versus technical corrections? and what are the patterns of evaluative reasoning employed by English course participants in collaborative writing tasks?

The Social Aspects of Learning to Write. The concept of students working together to enhance their writing is largely based on the Social Constructivist theory, notably Vygotsky's Zone of Proximal Development (ZPD). From this perspective learning is a social process in which pupils, via contact, are able to achieve a higher level of comprehension than would be possible individually. This means that when students work together, they provide each other a "scaffold" and combine their knowledge to tackle language issues (Storch, 2021) in an English writing class. Peer feedback is also emphasised by Hyland (2022) as it gives students an opportunity to see their writing from a different perspective. This change from writing for oneself to writing for a "audience" is a big step in becoming a better writer.

Characteristics of High-Quality Feedback

Recently, academics have gone beyond counting the number of comments students make and begun to look at the quality of remarks. "Feedback is only effective if it leads to meaningful changes in the writing and not just fixing a few commas," say Liu and Hansen (2022). To measure this quality, three different levels of participation might be considered (Yu & Hu 2021; Winstone et al. 2022):

- Level 1 (Surface/Corrective): These are comments that only point out small errors like spelling or tenses without explaining why they are wrong.
- Level 2 (Cognitive/Explanatory): Here, the student identifies a problem in the logic or content and gives a reason for their critique.
- Level 3 (Reflective/Strategic): This is the highest level. The student finds a problem, explains it, and then suggests a specific way to fix it.
- They have metalinguistic awareness in the sense that they are able to

communicate about language rules and strategies clearly when they reach Level 3 (Lee, 2023).

The Problem of Critical Reflection

We want students to give insightful, strategic input, yet sometimes it's not simple for students to give feedback. Critical reflection involves 'feedback literacy' which is the capacity to evaluate a piece of work against a specific standard and then convey that evaluation effectively (Winstone et al. , 2022). This is a problem for many EFL learners since they may not feel confident enough in their own English skills to "judge" anybody else. According to Lee (2023), students often make safe, easy compliments such as good job without sufficient experience or the correct words, just to get the assignment done.

Factors Influencing Students' Evaluative Hesitation

Feedback tends to stay at a surface level for social and mental reasons, too. Nguyen et al. (2022) explain that delivering feedback is actually rather exhausting for the brain a "cognitive overload." Students have to read and understand and evaluate at the same time, therefore they typically prefer to remedy the easy faults rather than deal with the difficult concerns.

In many cultures and especially in a mixed-age environment such as a private course, "politeness" is also a huge element. According to Brown and Levinson's (1987) Face Theory, individuals often avoid direct criticism in order to maintain interpersonal harmony and protect others' social face. Therefore, they may choose to be "nice" rather than honest and helpful. Teng (2022) warns that peer feedback can easily turn into a mechanical "checklist" activity without appropriate guidance from the teacher, with no actual learning taking place. We need to examine how to foster more critical and honest reflection in the classroom.

2. Method

Research Design

Qualitative Content Analysis (QCA) was used to analyse the depth and quality of students' peer feedback. This method was chosen because it allows for a detailed investigation of the evaluative reasoning of participants, rather than simply quantifying responses (Liu & Hansen, 2022).

Participants and Context

Fifteen students taking a private English course participated in the study. The class was structured in a mixed-age format, with learners from a variety of backgrounds and age groups. This context was intentionally selected because mixed-age interaction creates distinctive evaluative dynamics related to politeness, confidence, and social hierarchy during peer-feedback activities. Participants were all at an intermediate level of English. Convenience sampling was employed due to availability and the natural classroom setting.

Research Instrument

Data collection involved the use of a Peer Feedback Guideline and Worksheet. An evaluative rubric adapted from Winstone et al. (2022) and Hyland (2022) was employed to classify the feedback into three levels: corrective (Level 1), explanatory (Level 2), and strategic (Level 3).

Category	Description	Level
Corrective	Focuses on grammar/spelling with no explanation.	Level 1 (Surface)
Explanatory	Finds a problem and gives a logical reason.	Level 2 (Cognitive)
Strategic	Evaluates the quality and suggests a way to fix it.	Level 3 (Reflective)

Comments at Level 3 are considered "High Evaluative Quality."

Additionally, a short reflective survey was administered to explore students' reasoning behind their feedback. The reflective survey consisted of short open-ended questions exploring students' perceptions, emotional discomfort, and challenges experienced during peer-feedback activities.

Procedure

The procedure began with an introduction to the distinction between 'surface-level' and 'deep' feedback. Students wrote an argumentative essay in pairs. Each participant provided at least three comments to their partner using the worksheet. Finally, participants answered reflective questions regarding their experiences and perceived difficulties during the activity.

Data Analysis

All peer comments were coded to identify the distribution of types of feedback and answer the first research question (RQ1) according to the three-level evaluative rubric. During the first stage, comments were categorized according to the three evaluative levels. In the second stage, similar codes were grouped into broader themes such as politeness strategies, hesitation, technical focus, and life-experience reasoning. Patterns of evaluative reasoning were identified in the qualitative analysis of the notes from coding to answer the second research question (RQ2). To ensure reliability, a colleague independently reviewed approximately 30% of the coded data to confirm coding consistency. The coding agreement reached approximately 87%, and disagreements were discussed until consensus was achieved.

3. Result

The analysis of data was designed to evaluate the quality of 15 extracts of peer feedback received from the participants. To address the first research question (RQ1), the comments were coded for the amount of evaluative engagement. Table 1 shows the distribution of these levels and provides an overview of how students prioritised their responses.

Table 1. Distribution of Peer-Comment Quality Levels

Category	Level of Evaluative Quality	Frequency (n=15)	Percentage (%)
Corrective	Level 1 (Surface-level)	9	60%
Explanatory	Level 2 (Cognitive-level)	4	26.7%

Strategic	Level 3 (Reflective-level)	2	13.3%
Total		15	100%

Table 1 shows a high trend for surface level interaction. The majority of comments (60%) were identified as Level 1, where students focused almost completely on rectifying technical mistakes such as spelling, punctuation and simple grammar. The results show that just a tiny percentage of the feedback was provided at Level 3 (13.3%), indicating that it is significantly more difficult for learners to provide strategic solutions than to identify problems. This distribution indicates that while students are involved in the feedback process, they are more focused on “cleaning up” the text rather than critically evaluating its academic substance. To further address RQ2, we analysed many example extracts to discover patterns of evaluative reasoning. The results demonstrate major differences among the participants, mostly with age and conversational style. The younger participants were more prone to make safer, surface-level comments or to use cautious language when making suggestions. In contrast, elder participants showed a higher focus on content and connection to the real world, although their input was not always accompanied by obvious linguistic answers. These results suggest that evaluative quality is a function of both linguistic skill and social and experience characteristics.

Table 2. Samples of Student Peer Comments and Coding Results

Extract No.	Participant (Age)	Original (Verbatim)	Comment	Coding Level	Researcher's Patterns Found	Notes /
1	Student (17)	3	"Check your spelling in paragraph 2, 'tehnology' is wrong."	Level 1	Technical Focus: purely mechanical; no engagement with content.	
2	Student (19)	7	"I think maybe... if you don't mind, the intro is a bit long?"	Level 2	Politeness Pattern: Identifying a problem but using heavy "hedging" (hesitation).	
3	Student (42)	9	"This argument about social media is too simple for real world."	Level 2	Life Experience Logic: Engaging with the depth of the idea, but lacking a linguistic solution.	
4	Student (24)	5	"The transition is jumpy. You should add a bridge sentence to	Level 3	Strategic Awareness: High metalinguistic skill; providing a clear fix.	

connect them."					
5	Student 12 (35)	"This sentence is confusing. Better use 'however' to show contrast."	Level 3	Problem-Solution: Identifying a logical flaw and providing a specific vocabulary tip to improve clarity.	

The differences in evaluative reasoning tendencies among age groups are demonstrated in Table 2. Younger participants like Student 3 and Student 7 were more likely to either focus on safe surface-level corrections or use very careful language when advocating change. In contrast, the older participants (e.g., Student 9 and Student 12) seemed to draw on their life experiences in judgement, giving more emphasis to the “logic” and “real-world” significance of the arguments. Qualitative data reveals that while language proficiency is vital, social maturity and confidence also play a big impact in how a student chooses to assess their peer’s work. The findings also suggest that argumentative writing tasks may require stronger instructional scaffolding to encourage deeper evaluative engagement beyond technical correction.

A clearer comparison between age groups reveals that younger participants (17–24 years old) produced most Level 1 comments and frequently used hedging expressions when offering criticism. In contrast, older participants (30–45 years old) demonstrated stronger engagement with argument quality and real-world logic, although they did not always provide explicit linguistic solutions.

The peer feedback excerpts provided supplementary insight into students’ evaluative behaviour in addition to the reflective survey data. Several individuals were cautious to provide critical feedback especially when providing input to elder peers. For example, one participant said, “I was afraid my comment might sound rude if I corrected someone older than me”. Others said they had difficulty expressing their ideas because they did not have the words to do so. For example, one participant said, “I know something is wrong, but I do not know how to describe it in English.” These replies corroborate the findings that the degree of evaluative engagement is affected by both social circumstances and linguistic constraints.

4. Discussion

The findings of this study show a considerable discrepancy between the students’ abilities in identifying surface-level faults and their ability for deep critical thinking. To address RQ1, it is evident that the peer comments are mainly technical adjustments (Level 1) and do not involve critical engagement. Students were good at catching “local” difficulties, e.g., typos or wrong tenses, but they rarely went into “global” issues such as the power of the argument or the logical flow of the essay.

The dominance of Level 1 feedback accords with the concept of “feedback literacy” examined by Winstone et al. (2022). Most of the participants, especially those in the younger age group (17-19 years old), seemed to approach the peer feedback job as a proofreading task. For example, a 17-year-old student (Student 3) devoted exclusive attention on spelling: “You missed the ‘s’ in ‘technologies.’” Many EFL learners consider linguistic accuracy the primary goal of writing, frequently to the detriment of communicative depth. This conclusion is similar with other studies (e.g., Yu & Hu, 2021), which also indicated that surface-level input was dominant in the EFL environments.

However, when we look at RQ2, the evaluative reasoning processes, the mixed-age course dynamics produced particularly significant findings. The feedback exchanges showed a clear ‘Social Hierarchy Pattern’. For example, a 19-year-old student giving criticism to a 40-year-old professional was likely to be too careful, employing significant “hedging” or politeness methods. One comment read: “I suppose maybe... if you don't mind, the introduction is a touch long maybe?” This supports This finding aligns with Brown and Levinson’s (1987) concept of face-saving behaviour, in which individuals attempt to maintain interpersonal harmony and avoid potentially threatening social interaction. In the Indonesian culture, where seniority in age is greatly respected, it is generally culturally unacceptable or “impolite” for younger pupils to give a Level 3 (Strategic) critique to an older colleague. The findings reflect earlier research indicating social and cultural norms play an important impact in the feedback practices in EFL classrooms.

This phenomenon can also be interpreted through Brown and Levinson’s Face Theory, which explains how individuals attempt to maintain social harmony and avoid face-threatening acts during interaction. Younger participants appeared reluctant to provide direct criticism because evaluative comments toward older peers could potentially create interpersonal discomfort.

However, the older subjects (aged 30-45) exhibited a different pattern which we here call the ‘Life Experience Logic’ pattern. Their English grammar might not have always been great, but their logic (Level 2) was generally more grown up. “This argument about social media is too simple because in the real world it’s more complicated than that,” said Student 9, a 42-year-old participant. Although this participant was not able to deliver a linguistic answer (Level 3), their critical engagement with the subject was significantly stronger than that of the younger students. This aligns with Lee’s (2023) statement that evaluative quality is not just about linguistic proficiency but also about the reviewer’s grasp of the text’s goal.

This also shows “cognitive overload” across all age groups due to the paucity of Level 3 (Strategic) remarks, as explained in Cognitive Load Theory (Sweller, 2011). Participants reported feeling “tired” attempting to understand their peer’s views while also checking for grammar. This mental strain of course led them to the “easiest” way to rectify an error. To convert students from “correctors” to

“evaluators” it is not enough to supply them with a rubric. Rather we need to provide students with the unique “metalinguistic vocabulary” to be able to provide answers without feeling socially awkward.

The findings indicate that teacher scaffolding is essential in promoting higher-order peer feedback. Explicit modelling of constructive criticism, sentence starters, and guided peer-review activities may reduce evaluative hesitation and help students focus more on idea development and argumentative quality.

Compared with same-age peer-feedback environments discussed in previous studies, mixed-age classrooms appear to involve stronger sociocultural sensitivity related to respect, hierarchy, and interpersonal politeness. These conditions may intensify students’ tendency to avoid direct criticism.

In summary, the evaluative quality of this mixed-age English course is mainly formed by language constraints and social-cultural considerations. The propensity to “adopt low-risk evaluative strategies” with technical adjustments distracts students from the deep introspection needed for advancement in academic writing. The results imply that future educational interventions should place greater emphasis on the development of students’ “social-emotional feedback literacy” in order to enable them to overcome age-related barriers and concentrate on the quality of the argument.

5. Conclusion

The evaluative quality of peer comments in a mixed-age English class was examined in the present study. The results suggest that in response to RQ1, most participants still focus on surface-level technical fixes (Level 1) rather than critical reflection (Level 3). A “feedback gap” is evident when students tend to comment on the language and spelling rather than the power of the argument.

Regarding RQ2, the patterns of evaluative reasoning are subject to the ages and social dynamics of the participants. Younger students are more polite in their criticism to older students, whereas older students are more in-depth in their logical engagement but lack the linguistic abilities to give strategic solutions. This means that the two main barriers to good quality peer feedback in a non-formal, mixed-age context are ‘politeness’ and ‘cognitive overload’.

However, this study has some drawbacks, despite its merits. First, the sample size was quite small ($n = 15$) which could restrict generalisability of the findings. Second, the findings were obtained from a single private English course, which may not be representative of broader EFL classroom settings. Third, the study was limited mainly to written peer comments and did not include in-depth interviews that might have provided more in-depth insights into students’ thinking.

Future study should include bigger and more diverse populations of participants as well as other data sources, such as interviews or classroom observations to acquire a more complete picture of the evaluating methods within peer feedback.

It is recommended that teachers do not presume that students know how to give useful feedback. Teachers should provide students with sentence starters and

teach them how to provide constructive suggestions. In mixed-age classrooms, it is important to create a safe climate in which all are recognised as equal “writers” so that more honest reflections and strategic reflections are encouraged.

7. References

- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Cho, K., & MacArthur, D. (2020). Learning by reviewing: The impact of peer feedback on the reviewer’s own writing. *Journal of Educational Psychology*, 112(1), 22-35.
- Hyland, K. (2022). *Teaching and Researching Writing* (4th ed.). Routledge.
- Ismayanti, D., Said, Y. R., Usman, N., & Nur, M. I. (2024). The Students Ability in Translating Newspaper Headlines into English A Case Study. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 12(1), 108-131.
- Lee, I. (2023). *L2 Writing Teacher Education: Issues and Challenges*. Springer Nature.
- Liu, J., & Hansen, J. G. (2022). *Peer Response in Second Language Writing Classrooms*. University of Michigan Press.
- Masruddin, M., & Nasriandi, N. (2022). Lexical and Syntactical Errors Performed by Junior High School Student in Writing Descriptive Text. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 10(1), 1094-1100.
- Mercer, S., & Dörnyei, Z. (2020). *Engaging Language Learners in Contemporary Classrooms*. Cambridge University Press.
- Min, H. T. (2016). Training students to become better peer reviewers. *ELT Journal*, 70(1), 75-85.
- Rahimi, M. (2021). The role of teacher scaffolding in peer feedback quality. *International Journal of Applied Linguistics*, 31(2), 245-260.
- Storch, N. (2021). Collaborative writing in L2 classrooms: Why and how. *Journal of Second Language Writing*, 52, 100-112.
- Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, 55, 37-76.
- Teng, F. (2022). *Classroom Writing Assessment and Feedback in L2 Contexts*. Springer.
- Topping, K. J. (2021). *Peer Assessment: Learning from Reviewing*. Routledge.
- Villamil, O. S., & de Guerrero, M. C. (2020). Peer revision in the L2 classroom: Social and cognitive dimensions. *Applied Linguistics*, 41(4), 512-528.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2022). Feedback literacy: A systematic review and model for learner engagement. *Assessment & Evaluation in Higher Education*, 47(3), 320-335.
- Yu, S., & Hu, G. (2021). Understanding the evaluative quality of peer feedback in EFL writing. *Assessing Writing*, 48, 100-114.