

Exploring EFL Students' Peer Evaluation Practices in Collaborative Projects: Patterns, Biases, and Pedagogical Implications

Lutfi Istikharoh

Universitas Muhammadiyah Purwokerto

Corresponding Author: Lutfi Istikharoh, lutfiistikharoh@ump.ac.id

ARTICLE INFO

Keywords: Peer Evaluation, Group Work, EFL Students, TEFL Classroom, Assessment Literacy

Received : 28, June

Revised : 10, July

Accepted: 12, August

©2025 Istikharoh: This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

This study investigated how EFL students in a teacher education program engage in peer evaluation during collaborative group projects. Using a structured rubric across five dimensions—ideas, responsibility, feedback, cooperation, and participation—591 peer ratings were analyzed to reveal patterns in evaluative behavior. The results indicate a tendency toward leniency and uniform high scores, driven by sociocultural expectations of harmony and positive bias, including halo effects. Variability in scoring and limited qualitative justification suggest a need for improved assessment literacy and rubric interpretation. The discussion highlights the influence of cultural norms and the limitations of student evaluative skills, arguing that peer assessment should be supported by structured training and reflective practice. The study concludes that while peer evaluation holds promise for promoting critical thinking and accountability in TEFL settings, its educational value depends heavily on how it is integrated into pedagogy. Implications are drawn for curriculum design, instructional scaffolding, and the professional development of pre-service English teachers.

INTRODUCTION

Peer evaluation has increasingly emerged as a pedagogical strategy in higher education, especially within collaborative and constructivist learning frameworks. It aligns with the paradigm shift toward student-centered learning, which emphasizes autonomy, mutual accountability, and reflective practice. In collaborative tasks, peer evaluation enables students not only to assume ownership over their own learning but also to participate in the assessment of their group members' contributions. This process is argued to foster deeper cognitive engagement, social responsibility, and evaluative judgment (Li & Gao, 2015).

In the context of Teaching English as a Foreign Language (TEFL), group projects are widely integrated into instructional practices to promote interaction, communicative competence, and cooperative learning. As prospective teachers, EFL students are often expected to engage in reflective collaboration, where peer feedback and evaluation become both a learning process and a tool for professional development. When applied effectively, peer assessment can support the development of students' understanding of performance criteria, enhance their critical reflection, and cultivate metacognitive awareness (Farahian & Ebadi, 2023).

Despite its potential, peer evaluation presents complex challenges. Students may struggle with subjectivity, rating bias, and discomfort in evaluating their friends critically. Cultural norms – especially in collectivist societies such as Indonesia – may further discourage honest appraisal due to a strong emphasis on social harmony (Rahmatillah & Fajrita, 2022). In addition, several studies suggest that without sufficient training and assessment literacy, students often find it difficult to interpret rubrics, provide consistent ratings, or differentiate between various dimensions of performance (Ali Salami & Mastour Alharthi, 2022; Prastikawati et al., 2024; Sohrabi et al., 2022; Thị & Lộc, 2016). These issues may compromise the reliability and pedagogical value of peer evaluation in practice.

Although numerous studies have explored the theoretical and practical dimensions of peer assessment, there is still a limited body of empirical research examining how EFL students actually implement evaluation criteria in real classroom settings, particularly in Asian or collectivist cultural contexts (Li & Gao, 2015; Wardani et al., 2021). Prior research has largely focused on students' perceptions or the alignment between peer and teacher scores, leaving a gap in understanding the evaluative behavior, bias patterns, and rubric application practices among EFL learners working on collaborative group tasks. Moreover, studies that combine quantitative rating patterns with qualitative feedback analysis are still relatively rare in TEFL literature, which limits the comprehensiveness of our insights into how students navigate the complexities of peer evaluation.

This study aims to address these gaps by investigating student-led peer evaluation practices in a TEFL course at an Indonesian university. Specifically, the study explores how pre-service EFL teachers assess their peers during group projects, what rubric dimensions they prioritize, and what behavioral or

cognitive tendencies are reflected in their scoring patterns. Through a mixed-methods analysis of rubric-based quantitative data and available qualitative comments, this research offers a grounded perspective on how peer evaluation is enacted in practice – highlighting both its affordances and limitations.

The following research questions guide this study:

1. How do EFL students evaluate their peers during collaborative group projects?
2. What evaluation criteria do students prioritize in assessing their group members?
3. What patterns or challenges emerge from students' evaluation scores and comments?

By addressing these questions, the study seeks to advance a more nuanced understanding of peer evaluation as a pedagogical practice in EFL teacher education. It aims to illuminate how learners negotiate roles as both evaluators and collaborators, how sociocultural factors influence scoring behaviors, and how institutional assessment systems might be refined to support the development of more valid, equitable, and educationally meaningful peer evaluation practices. The novelty of this research lies in its integration of cultural, pedagogical, and empirical insights into peer evaluation behavior – particularly in an underrepresented context where the interplay between collaboration, assessment, and sociocultural dynamics remains insufficiently explored.

LITERATURE REVIEW

Peer Assessment in EFL Context

Peer assessment refers to a process where students evaluate the performance or contributions of their peers using established criteria, often within collaborative or project-based learning environments (Topping, 1998). In the EFL context, peer assessment serves not only as an evaluative mechanism but also as a pedagogical tool that fosters learner autonomy, critical thinking, and language awareness (El-Mowafy, 2010; Saito & Fujita, 2009). Several studies have shown that when students participate in peer assessment, they tend to become more conscious of quality standards, actively engage in reflective thinking, and internalize evaluation criteria (El-Mowafy, 2010; Steensels et al., 2006).

In group project settings, particularly within TEFL courses, peer evaluation offers opportunities for learners to assess not only linguistic output but also interpersonal and collaborative behaviors such as responsibility, initiative, and cooperation. This aligns with the communicative and interactive nature of language learning, wherein student-led evaluation can enhance both social and academic outcomes (Alqarni & Alshakhi, 2021; El-Mowafy, 2010; Kizil, 2019).

Student Judgement and Assessment Behavior

While peer assessment promotes student agency, it also raises concerns regarding the fairness and reliability of peer judgments. Research has found that students often struggle to apply criteria consistently, and their evaluations may be influenced by social relationships, perceived effort, or personal bias (Panadero

& Jonsson, 2020). In some cases, peer evaluation may result in overly generous scores to avoid interpersonal conflict, or conversely, harsh assessments due to group tension (Zhao, 2010).

Moreover, students' understanding of evaluation criteria – also known as assessment literacy – plays a crucial role in the effectiveness of peer evaluation (Tai et al., 2018). Without adequate guidance and training, students may misinterpret descriptors, rely on subjective impressions, or mirror dominant group dynamics rather than offer objective judgments.

Collaborative Learning and Group Dynamic in TEFL

Collaborative group projects are a central feature in TEFL pedagogy, often designed to simulate authentic language use and promote cooperative learning (Richards & Rodgers, 2014). In such settings, evaluation of peer performance is inherently tied to the group process. Studies in EFL classrooms have highlighted that group dynamics – including leadership, participation equity, and task division – affect how students perceive and assess one another (Barkley et al., 2005; Cohen & Lotan, 2014).

Effective peer assessment must therefore account for these dynamics and be supported by transparent rubrics, ongoing instructor facilitation, and reflective opportunities for students. When peer evaluation is embedded in a well-structured learning design, it can reinforce communication skills, responsibility-sharing, and metacognitive awareness (Michaels & O'Connor, 2015).

Validity, Reliability, and Challenges in Peer Evaluation

The validity and reliability of peer-assigned scores have long been debated. Studies show that while peer ratings can align reasonably well with instructor judgments, variability remains high, particularly in subjective areas such as attitude or initiative (Falchikov & Boud, 1989). To address this, researchers emphasize the need for clear rubric design, calibration activities, and training in giving constructive feedback.

Further, peer evaluation can serve as a formative learning tool rather than merely a summative one. Emphasis on process-oriented reflection, rather than final grades, helps students develop evaluative competence and ownership of learning (Sadler, 1989). Such perspectives are especially important in EFL settings, where students are developing both language skills and academic behaviors simultaneously.

These challenges in ensuring the validity and reliability of peer-assigned scores are closely related to the conceptual framework illustrated in Figure 1. The framework begins with the premise that peer evaluation, when integrated into collaborative learning, relies heavily on rubric-based assessment to guide student judgment. However, the effectiveness of this rubric use is contingent upon the students' assessment literacy – their ability to interpret and apply criteria meaningfully. Assessment literacy influences two key practices: how students utilize rubrics and how they rate criteria. Deficiencies in either domain may lead to evaluation biases.

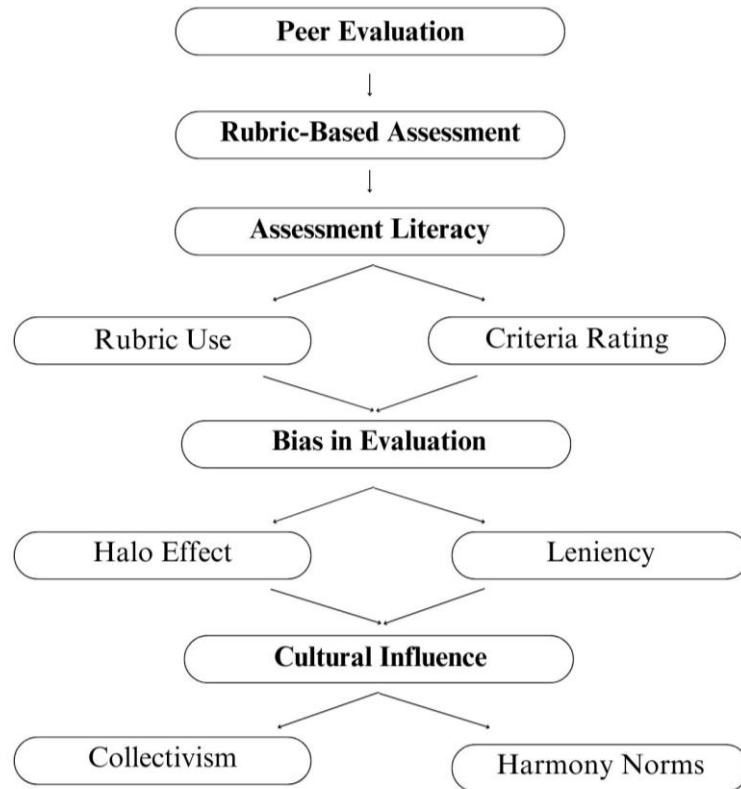


Figure 1. Conceptual Framework (images must be in good quality)

As depicted in the diagram, bias in evaluation can take the form of the halo effect—where a single positive trait colors overall judgment—or leniency, where students hesitate to assign low scores due to personal relationships or social discomfort. These biases are further shaped by broader cultural influences, particularly in collectivist settings like Indonesia. Cultural norms emphasizing group harmony and avoidance of conflict often discourage students from giving critical feedback, even when it is warranted. Hence, the framework contextualizes peer evaluation not only as a pedagogical tool but also as a socio-culturally situated practice, where cognitive, interpersonal, and cultural dimensions converge to shape student judgment. This interplay highlights the need for more structured support, such as rubric training and reflective guidance, to foster valid, fair, and educationally meaningful peer evaluation in EFL classrooms.

METHODOLOGY

Research Design

This study employed a descriptive quantitative research design with embedded qualitative elements. The primary aim was to analyze student-generated peer evaluation data collected from a group project in a TEFL (Teaching English as a Foreign Language) course. Quantitative data were derived from rubric-based scoring, while qualitative insights were explored through students' written comments where available. This mixed approach allowed for a comprehensive understanding of students' evaluative behavior in collaborative language learning contexts.

Context and Participants

The research was conducted in the TEFL A course at an English Language Education undergraduate program in Central Java, Indonesia. The course incorporated a project-based assignment in which students worked in small groups and evaluated their peers at the end of the collaboration process. A total of 61 active students participated in the peer evaluation activity, grouped into several project teams. The participants were pre-service EFL teachers in their fifth semester, all of whom had prior experience with group projects and were familiar with basic peer assessment procedures.

Instrumentation

The primary instrument employed in this study was a peer evaluation rubric designed to assess multiple dimensions of student contribution in collaborative group work. The rubric encompassed five key criteria: responsibility, initiative, cooperation, creativity, and participation. Each of these components was selected to capture both individual accountability and group dynamics, reflecting cognitive, behavioral, and interpersonal aspects of teamwork. The structured rubric provided students with a clear framework for evaluating their peers and served as the basis for the data analysis in this research.

Each criterion was rated on a scale of 1 to 4, where 1 indicated minimal performance and 4 indicated excellent performance. Students were instructed to evaluate each of their group members (excluding themselves) based on the rubric. The evaluation form also included an optional comment section to justify or elaborate on the scores given.

The rubric was designed to capture key dimensions of group participation and was introduced in class prior to its application, ensuring that students were aware of the criteria and how to apply them.

Data Collection and Analysis

Peer evaluation data were collected at the conclusion of the group project period. Each student submitted a completed evaluation form, resulting in a dataset of multiple peer-generated scores across different teams. The quantitative data were compiled in Microsoft Excel for analysis.

The analysis process involved the following steps:

1. Calculating the mean score per criterion for each student.
2. Identifying the highest and lowest scoring criteria across the cohort.
3. Examining the consistency of scores within and between groups.

Data visualization tools (e.g., tables and bar charts) were used to support the interpretation of scoring trends and facilitate comparison across groups.

RESEARCH RESULTS

Quantitative Analysis of Peer Evaluation Scores

A total of 591 peer evaluation entries were collected from 36 student evaluators, involving assessments of 68 different peers across five criteria: ideas/initiative, responsibility, feedback, cooperation, and participation. Each evaluator rated multiple group members using a 4-point rubric, where 1

indicated minimal contribution and 4 indicated excellent performance. Descriptive statistics of the scores are presented in Table 1.

Table 1. Descriptive Statistics by Criterion

Criterion	n (data)	Mean	Std Dev	Min	Q1 (25%)	Median (50%)	Q3 (75%)	Max
Ideas	153	3.05	1.03	1.00	3.00	3.00	4.00	4.00
Responsibility	134	2.96	1.01	1.00	2.25	3.00	4.00	4.00
Feedback	108	3.05	1.02	1.00	3.00	3.00	4.00	4.00
Cooperation	72	3.17	0.96	1.00	3.00	3.00	4.00	4.00
Participation	57	3.14	1.01	1.00	2.00	3.00	4.00	4.00

As shown in Table 1, all five criteria received mean scores between 2.96 and 3.17, indicating generally positive peer evaluations. Among these, Cooperation (M = 3.17) and Participation (M = 3.14) were rated the highest, suggesting that students strongly recognized collaborative effort and consistent involvement in group tasks.

Conversely, Responsibility (M = 2.96) received the lowest mean score, potentially indicating variability in how students fulfilled their assigned duties within the team. Notably, the standard deviations for all criteria range from 0.96 to 1.03, reflecting moderate dispersion and suggesting that peer judgments were not entirely uniform.

Quartile distribution further shows that the majority of scores clustered around 3 and 4, aligning with the general tendency of students to rate their peers favorably. However, the presence of minimum scores as low as 1.0 across all criteria also reveals that some group members were perceived as underperforming or disengaged, at least in isolated cases.

These findings highlight overall positive peer perceptions while also hinting at subtle differences in how students interpret and apply each evaluative criterion. A deeper examination of scoring patterns and potential challenges will be discussed in the next section.

Patterns and Challenges in Peer Evaluation Behavior

The distribution of scores across peer evaluation criteria reveals several consistent patterns that reflect both the strengths and the limitations of student-led assessment in group settings.

Leniency and Positive Bias

Across all five criteria, the majority of peer scores fell between 3 and 4, with mean values above 2.9 and medians consistently at 3.0. This suggests a pattern of lenient evaluation, where students are inclined to assess their peers positively – possibly to maintain social harmony or avoid confrontation. The relatively low occurrence of scores below 2.0 indicates that critical judgment was rarely exercised, even in cases where performance may have been lacking.

This tendency aligns with findings from Panadero and Brown (2017), who highlight that in peer assessment contexts, social relationships and peer

pressure often inhibit honest, critical evaluations. As such, while the data reflect surface-level equity, they may obscure real disparities in effort or quality of contribution.

Variability in Judgment Across Criteria

While some criteria (such as Cooperation and Participation) received higher overall scores, others—particularly Responsibility and Feedback—displayed more varied scoring. Responsibility showed a slightly lower mean (2.96) and greater dispersion, suggesting that students were more comfortable recognizing differences in task fulfillment. Feedback, which requires both interpersonal communication and metacognitive reflection, also had a slightly wider range of scores, indicating uncertainty in evaluating abstract behaviors like giving or receiving input.

These trends raise concerns about assessment literacy, or students' understanding of how to interpret and apply evaluation criteria (Tai et al., 2018). Without explicit training, students may struggle to distinguish between different forms of contribution, particularly those less visible or tangible, such as emotional support or idea generation.

Disparities in Rating Opportunities

The number of valid scores differed significantly across criteria—e.g., Ideas (n = 153) had far more entries than Participation (n = 57). This disparity likely stems from inconsistencies in how thoroughly students completed the evaluation forms, or from variations in how many group members were assessed per evaluator. This finding points to a procedural weakness in the data collection process: not all evaluators completed the full rubric for each peer, or some criteria may have been skipped due to lack of clarity or relevance.

Potential Halo Effects

Given the general tendency to score peers highly across all dimensions, it is possible that some students engaged in halo rating—assigning uniformly high scores based on a single positive impression (e.g., likability or dominance). This undermines the objectivity of the evaluation process and limits the ability to distinguish nuanced contributions within the group.

Limitations from Absence of Qualitative Commentary

Although the evaluation form included an open-ended comment section, few students provided meaningful written feedback. This limited the possibility of triangulating numerical scores with justifications or reflective insights. The lack of qualitative data may indicate either a reluctance to critique peers, or a lack of confidence in articulating feedback constructively.

DISCUSSION

This study investigated how EFL learners in a TEFL course engaged in peer evaluation of collaborative group project using a structured rubric comprising five key criteria. The analysis of 591 evaluation entries provided insights into the cognitive, cultural, and pedagogical dynamics that shaped the

peer assessment process. While the results revealed general patterns of successful engagement, they also highlighted limitations in evaluative depth, rating consistency, and assessment literacy.

The predominance of high scores across all evaluation criteria reflects a strong inclination toward lenient assessment. This trend is particularly evident in cooperation and participation, which received the highest mean scores with minimal dispersion. Such behavior appears to be culturally grounded; in collectivist contexts like Indonesia, students may be reluctant to critically assess peers due to the potential for interpersonal conflict or loss of group chemistry. Peer evaluation, in such contexts, may be guided less by objective performance differentiation and more by relational harmony. Previous research has consistently observed similar tendencies in Asian EFL settings, where social expectations of mutual respect can override individual accountability (Hapsari & Fatmasari, 2022; Hofstede, 2001). The inclination to maintain peer solidarity may explain why scores were generally skewed positively and why critical feedback was minimal or absent.

This tendency toward uniform high scoring also reflects the influence of the halo effect in the evaluation process. Students often assessed all dimensions of peer performance similarly, suggesting that one favorable trait—such as charisma, fluency, or leadership—might have shaped their overall perception. The presence of such bias undermines the diagnostic capacity of peer evaluation. Instead of reflecting differentiated contributions, the scores might represent generalized impressions. This aligns with psychological research on evaluative bias, particularly in peer-assessed settings without moderation or training (Panadero & Jonsson, 2020)). In this study, the lack of variation within evaluator scores and the high occurrence of full scores across categories support the claim that some students may have assessed based on interpersonal preference rather than performance per criterion.

Moreover, the data reveal challenges related to assessment literacy and rubric interpretation. Although a structured rubric was used, students appeared to interpret certain criteria inconsistently. Responsibility, for instance, exhibited greater variation in scores and received the lowest mean rating, suggesting uncertainty about what constitutes task fulfillment in a group context. Feedback and ideas, which require inferential judgment and reflective awareness, also displayed relatively more scoring divergence. These patterns indicate that students may lack the evaluative vocabulary and cognitive clarity needed to make precise judgments. This finding supports arguments by Tai et al. (2018) and (Sadler, 1989), who emphasize that effective assessment requires explicit instruction and experience. Without a clear understanding of how to evaluate abstract dimensions like feedback quality or initiative, students may default to vague or non-discriminating ratings.

The irregularity in the number of responses across criteria adds another layer of concern. While some criteria were assessed more than 150 times, others received fewer than 60 evaluations. This disparity may have resulted from confusion, fatigue, or selective completion. It also raises questions about procedural consistency in the data collection process. Such inconsistency

threatens the reliability of the findings, as the scoring system assumes equal attention to each rubric item. The problem is compounded by the lack of written comments accompanying most evaluations, which limits the researcher's ability to triangulate numerical scores with qualitative justification. This aligns with previous studies, such as those by Cartney (2010) that have noted how incomplete peer evaluation responses often reflect a lack of seriousness or confidence among student evaluators, particularly when feedback is not anonymous or scaffolded.

Despite these challenges, the findings affirm the pedagogical potential of peer evaluation, especially in teacher education settings. The act of evaluating peers can promote metacognitive growth, empathy, and accountability—skills essential for prospective educators. However, to realize this potential, peer assessment should not be treated as a stand-alone activity. It requires structured scaffolding that includes rubric explanation, rating practice, peer moderation, and self-reflection. Instructors should also consider incorporating mechanisms such as justification prompts and anonymous digital platforms to reduce social pressure and improve rating honesty (Nicol & MacFarlane-Dick, 2006; Zhao, 2010). Peer evaluation must be situated within a broader pedagogical design that cultivates assessment literacy and critical awareness among learners. Without this support, peer assessments may only replicate superficial judgments rather than foster transformative learning.

In summary, this study reveals both the strengths and vulnerabilities of peer evaluation in an EFL teacher education context. While students demonstrated high engagement and a willingness to assess, their evaluations were shaped by cultural, social, and pedagogical factors that influenced the objectivity and depth of their judgment. Addressing these factors through instructional design and evaluative training is essential if peer assessment is to contribute meaningfully to language teacher development.

CONCLUSIONS AND RECOMMENDATIONS

This study investigated the implementation of peer evaluation in collaborative group projects among EFL students in a teacher education program. By analyzing 591 peer assessment entries across five key rubric criteria—ideas, responsibility, feedback, cooperation, and participation—it reveals that while students are generally capable of completing peer evaluations, the quality, depth, and consistency of their assessments are shaped by multiple layers of cultural norms, interpersonal dynamics, and assessment literacy.

The results indicate that students tend to give favorable scores to their peers, often without significant differentiation among the criteria. This pattern suggests a prevailing culture of generosity, driven in part by a desire to maintain group cohesion and avoid interpersonal conflict. While this may contribute to a positive learning atmosphere, it simultaneously compromises the diagnostic value of the peer assessment tool. Furthermore, scoring behaviors were influenced by the halo effect, where a single strong impression about a peer colored judgments across unrelated criteria, potentially obscuring true differences in contribution.

Variations in response frequency and inconsistencies in scoring criteria also indicate procedural and cognitive limitations in how students engaged with the assessment task. These limitations underscore the importance of embedding assessment literacy into the curriculum. Students need structured opportunities to learn how to interpret rubrics, apply evaluative criteria fairly, and provide reasoned justification for their judgments. Without this pedagogical support, peer assessment risks becoming a symbolic or performative task rather than a meaningful learning activity.

From a broader perspective, this study highlights the dual role of peer assessment in TEFL contexts: it serves not only as a tool for accountability in group work, but also as a medium for cultivating reflective, responsible, and pedagogically aware future teachers. When designed and implemented thoughtfully, peer evaluation can develop metacognitive awareness and critical thinking – skills essential for professional development in language education.

Practitioners are therefore encouraged to treat peer evaluation not merely as an assessment method, but as an instructional opportunity. Incorporating guided training sessions, rubric calibration, anonymous submission platforms, and feedback reflection can significantly enhance both the reliability of peer ratings and the educational impact of the process. Future research could build on this study by exploring how different levels of scaffolding influence the validity and educational outcomes of peer evaluation, especially in non-Western or collectivist academic cultures.

In conclusion, peer evaluation remains a promising yet pedagogically delicate practice. Its success depends on how well students are prepared, guided, and supported in carrying out this evaluative responsibility. For teacher education programs, investing in peer assessment training is not merely about improving fairness in grading – it is about cultivating the habits of judgment, reflection, and professional accountability that define effective educators.

ADVANCED RESEARCH

While this study offers insights into peer evaluation practices among EFL students in collaborative projects, it is not without limitations. One key limitation lies in the reliance on student-generated rating data without consistent qualitative commentary. Although the rubric provided structured dimensions for assessment, the lack of written justification limited the ability to triangulate numeric scores with students' reasoning or reflections. Future research should incorporate mandatory reflective comments or think-aloud protocols to capture how students interpret and apply rubric criteria.

Another limitation pertains to the cultural and institutional specificity of the study. Conducted within a single Indonesian university context, the findings may reflect localized cultural norms – particularly the tendency toward social harmony and aversion to confrontation – which may not generalize to EFL classrooms in more individualistic societies. Comparative studies across cultural settings would deepen understanding of how collectivist or individualist orientations shape evaluative behaviors.

Additionally, this study focused primarily on the output of peer evaluations rather than the process. Future research could adopt a longitudinal or observational design to examine how students develop assessment literacy over time, how peer interactions shape their judgment, and how training interventions influence rating accuracy. Integrating qualitative interviews or focus group discussions would also provide richer insight into the emotional and cognitive dynamics underlying peer assessment decisions.

Finally, further research may explore technological integration in peer assessment, such as using anonymous digital platforms or AI-assisted tools, to reduce bias and increase rating validity. Investigating how such tools affect students' evaluative confidence, honesty, and critical engagement would be a valuable extension of this work.

In sum, this study lays a foundational understanding of how pre-service EFL teachers engage with peer evaluation in group tasks. To enhance both the rigor and reach of future studies, researchers are encouraged to adopt multimodal, comparative, and process-oriented approaches that address the social, cultural, and pedagogical complexity of peer assessment in diverse TEFL contexts.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to all colleagues and lecturers at the English Education Study Program of Universitas Muhammadiyah Purwokerto for their valuable insights, constructive feedback, and academic support throughout the research process. Special thanks are extended to the students who participated in this study for their honest engagement in the peer evaluation activity and their willingness to contribute to the research.

Appreciation is also conveyed to the university for the academic environment and administrative support that enabled the completion of this study. This research was supported in part by institutional funding provided by Universitas Muhammadiyah Purwokerto. The contribution of this financial support is gratefully acknowledged.

REFERENCES

- Ali Salami, F., & Mastour Alharthi, R. (2022). Improving Language Assessment Literacy for In-Service Saudi EFL Teachers. *Arab World English Journal*, 13(3), 536–554. <https://doi.org/10.24093/awej/vol13no3.35>
- Alqarni, T., & Alshakhi, A. (2021). The impact of negotiation as a social practice on efl writing peer assessment sessions. *Theory and Practice in Language Studies*, 11(10), 1334–1341. <https://doi.org/10.17507/tpls.1110.23>
- Barkley, E. F., Cross, K. P., & Major, C. H. (2005). *Collaborative Learning Techniques* (1st ed.). John Wiley & Sons, Inc.
- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment and Evaluation in Higher Education*, 35(5), 551–564. <https://doi.org/10.1080/02602931003632381>

- Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: strategies for the heterogeneous classroom*. Teachers College Press.
- El-Mowafy, A. (2010). Using peer assessment of fieldwork to enhance students' practical training. *Assessment & Evaluation in Higher Education*, 39(2), 223-241.
- Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, 59(4), 395-430. <https://doi.org/10.3102/00346543059004395>
- Farahian, M., & Ebadi, S. (2023). Collaborative digital writing and metacognitive knowledge in writing among TEFL students: the mediating role of online knowledge sharing. In *Technology Enhanced Learning* (Vol. 18).
- Hapsari, A., & Fatmasari, T. A. (2022). Online Self-Regulated Learning Strategies in the Process of Writing Undergraduate Thesis: A Survey Study. *Journal of English and Education (JEE)*, 8(2), 58-72. <https://doi.org/10.20885/jee.v8i2.24333>
- Hofstede, G. (2001). Culture's Recent Consequences: Using Dimension Scores in Theory and Research. *International Journal of Cross-Cultural Management*, 1(1), 11-30.
- Kizil, V. (2019). Peer Assessment in EFL Writing Classes. *The Journal of International Lingual Social and Educational Sciences*, 5(1), 101-111. <https://doi.org/10.34137/jilses.560099>
- Li, L., & Gao, F. (2015). Effect of Peer assessment on Project Performance of Students at Different Learning Levels. *Visual Communications and Technology Education Faculty Publications*, 33. https://scholarworks.bgsu.edu/vcte_pub/33
- Michaels, S., & O'Connor, C. (2015). Conceptualizing Talk Moves as Tools: Professional Development Approaches for Academically Productive Discussions. In *Socializing Intelligence Through Academic Talk and Dialogue* (pp. 347-361). American Educational Research Association. https://doi.org/10.3102/978-0-935302-43-1_27
- Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. In *Educational Research Review* (Vol. 30). Elsevier Ltd. <https://doi.org/10.1016/j.edurev.2020.100329>
- Prastikawati, E. F., Mujiyanto, Y., Saleh, M., & Fitriati, S. W. (2024). Teaching Practicum: Does It Impact to Pre-service EFL Teachers' Language Assessment Literacy? *World Journal of English Language*, 14(3), 242-251. <https://doi.org/10.5430/wjel.v14n3p242>
- Rahmatillah, R., & Fajrita, R. (2022). Journal of Social and Policy Issues Analyzing Factors Affecting the Effectiveness of Peer Assessment in EFL Teaching. *INDONESIAN JOURNAL OF TEACHING AND TEACHER EDUCATION*, 3(1). <https://doi.org/10.35308/xxxxx>

- Richards, J. C., & Rodgers, T. S. (2014). *APPROACHES AND METHODS IN LANGUAGE TEACHING* (3rd ed.). Cambridge University Press.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. In *Instructional Science* (Vol. 18).
- Saito, H., & Fujita, T. (2009). Peer-assessing peers' contribution to EFL group presentations. *RELC-Journal*, 40(2), 149–171. <https://doi.org/10.1177/0033688209105868>
- Sohrabi, Z., Ghanbari, N., & Abbasi, A. (2022). Iranian EFL teachers' perceptions of writing assessment literacy: a countrywide study. *Language Testing in Asia*, 12(1). <https://doi.org/10.1186/s40468-022-00172-7>
- Steensels, C., Leemans, L., Buelens, H., Laga, E., Lecoutere, A., Laekemen, G., & Simoens, S. (2006). Peer assessment: A valuable tool to differentiate between student contributions to group work? *Pharmacy Education*, 6(2), 111–118. <https://doi.org/10.1080/15602210600662279>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- Thị, N., & Lộc, H. (2016). *DEVELOPMENT OF VIETNAMESE PRE-SERVICE EFL TEACHERS' ASSESSMENT LITERACY*.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. <https://doi.org/10.3102/00346543068003249>
- Wardani, W. O., Sukyadi, D., & Purnawarman, P. (2021, April 16). Exploring Teacher Assessment Literacy in EFL Classroom. <https://doi.org/10.4108/eai.16-10-2020.2305253>
- Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, 15(1), 3–17. <https://doi.org/10.1016/j.asw.2010.01.002>